

AD-A130 799

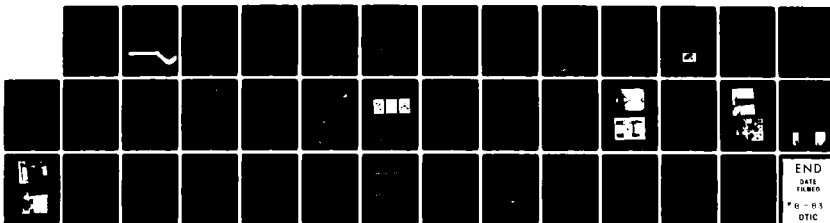
FACT RETRIEVAL FOR THE 1980'S(U) LAWRENCE LIVERMORE
NATIONAL LAB LIVERMORE CA V E HAMPEL JUL 81 UCRL-85749
AGARD-CPP-304 W-7405-ENG-48

1/1

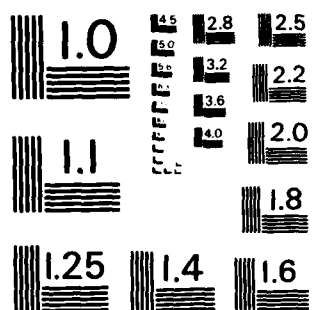
UNCLASSIFIED

F/G 5/2

NL



END
DATE
FILMED
* 6 - 83
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS - 1963 - A

1

UCRL - 85749
AGARD - CPP - 304

ADA130795

FACT RETRIEVAL FOR THE 1980s

Viktor E. Hampel

TRANSPORTATION SYSTEMS RESEARCH

This paper was published as Paper No. 6
in the proceedings of the Technical Information Panel
(TIP), Conference of the Advisory Group
for Aerospace Research and Development (AGARD),
North Atlantic Treaty Organization (NATO),
Munich, West-Germany,
September 9-10, 1981.

July 1981

DTIC FILE COPY

5-10-81
A

Lawrence
Livermore
Laboratory

Due to the fact that the information
contained herein is classified as
Secret, it is not to be released
without prior approval of the
Department of Defense.

FACT RETRIEVAL FOR THE 1980s

TABLE OF CONTENTS

SUMMARY	6-1
1. INTRODUCTION	6-1
2. THE STATUS OF FACT RETRIEVAL TODAY	6-2
Fact and Value	6-2
Datedness and Accuracy of Facts	6-3
The Cost of Facts	6-3
The Storage of Facts	6-4
The Analysis of Facts	6-5
Directories to Databases of Facts	6-6
The Tools of Fact Retrieval	6-6
The Marketing of Facts	6-7
Other Issues of Fact Retrieval	6-8
3. PROSPECTS FOR THE FUTURE	6-8
The Communication of Facts	6-8
Standards for Data Exchange	6-10
The International System of Units (SI) for Numeric Facts	6-10
Standards for Fact Attributes	6-10
The Copyrighting of Facts	6-11
The Electronic Office, Video Transmission, and Electronic Mail	6-12
Encoding and Decoding of Facts	6-13
The User Interface	6-13
Extraction of Intelligence	6-15
The Intelligent Gateway Computer	6-18
4. FACT RETRIEVAL IN THREE DIMENSIONS AS A FUNCTION OF TIME	6-20
Integral Holography	6-20
Stereo Vision of Facts	6-22
5. THE INTEGRATED "TECHNOLOGY INFORMATION SYSTEM" AT LLNL	6-26
Database Management Capabilities	6-26
Modeling	6-28
Communications	6-30
Distributed Networking	6-31
CONCLUSION	6-33
ACKNOWLEDGEMENT	6-33
REFERENCES	6-34

Distribution For
 1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593
 594
 595
 596
 597
 598
 599
 600
 601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611
 612
 613
 614
 615
 616
 617
 618
 619
 620
 621
 622
 623
 624
 625
 626
 627
 628
 629
 630
 631
 632
 633
 634
 635
 636
 637
 638
 639
 640
 641
 642
 643
 644
 645
 646
 647
 648
 649
 650
 651
 652
 653
 654
 655
 656
 657
 658
 659
 660
 661
 662
 663
 664
 665
 666
 667
 668
 669
 670
 671
 672
 673
 674
 675
 676
 677
 678
 679
 680
 681
 682
 683
 684
 685
 686
 687
 688
 689
 690
 691
 692
 693
 694
 695
 696
 697
 698
 699
 700
 701
 702
 703
 704
 705
 706
 707
 708
 709
 710
 711
 712
 713
 714
 715
 716
 717
 718
 719
 720
 721
 722
 723
 724
 725
 726
 727
 728
 729
 730
 731
 732
 733
 734
 735
 736
 737
 738
 739
 740
 741
 742
 743
 744
 745
 746
 747
 748
 749
 750
 751
 752
 753
 754
 755
 756
 757
 758
 759
 760
 761
 762
 763
 764
 765
 766
 767
 768
 769
 770
 771
 772
 773
 774
 775
 776
 777
 778
 779
 780
 781
 782
 783
 784
 785
 786
 787
 788
 789
 790
 791
 792
 793
 794
 795
 796
 797
 798
 799
 800
 801
 802
 803
 804
 805
 806
 807
 808
 809
 810
 811
 812
 813
 814
 815
 816
 817
 818
 819
 820
 821
 822
 823
 824
 825
 826
 827
 828
 829
 830
 831
 832
 833
 834
 835
 836
 837
 838
 839
 840
 841
 842
 843
 844
 845
 846
 847
 848
 849
 850
 851
 852
 853
 854
 855
 856
 857
 858
 859
 860
 861
 862
 863
 864
 865
 866
 867
 868
 869
 870
 871
 872
 873
 874
 875
 876
 877
 878
 879
 880
 881
 882
 883
 884
 885
 886
 887
 888
 889
 890
 891
 892
 893
 894
 895
 896
 897
 898
 899
 900
 901
 902
 903
 904
 905
 906
 907
 908
 909
 910
 911
 912
 913
 914
 915
 916
 917
 918
 919
 920
 921
 922
 923
 924
 925
 926
 927
 928
 929
 930
 931
 932
 933
 934
 935
 936
 937
 938
 939
 940
 941
 942
 943
 944
 945
 946
 947
 948
 949
 950
 951
 952
 953
 954
 955
 956
 957
 958
 959
 960
 961
 962
 963
 964
 965
 966
 967
 968
 969
 970
 971
 972
 973
 974
 975
 976
 977
 978
 979
 980
 981
 982
 983
 984
 985
 986
 987
 988
 989
 990
 991
 992
 993
 994
 995
 996
 997
 998
 999
 1000



FACT RETRIEVAL IN THE 1980s

Viktor E. Hampel
Technology Information System
Transportation Systems Research
Lawrence Livermore National Laboratory
P.O. Box 808, Livermore, California 94550, USA

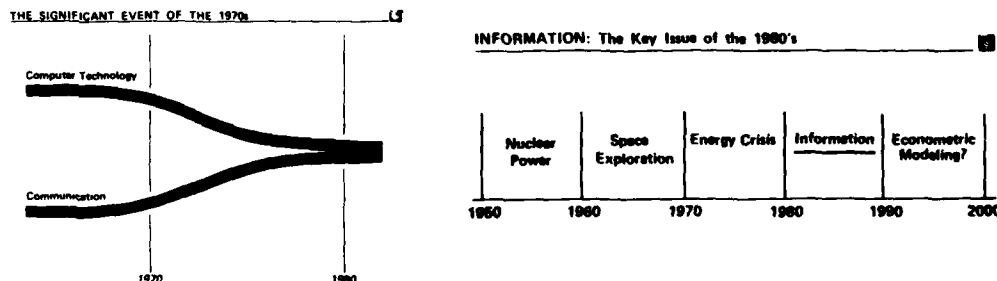
SUMMARY

This report reviews prevailing methodologies of fact retrieval in science and technology and makes surprise-free projections for the decade to come: Numeric databases are shown to overtake in size and number the large bibliographic collections. This is expected to lead toward more sophisticated, interactive data analysis techniques with graphical display options. The availability of low-cost intelligent computer terminals, micro- and minicomputers, is shown to make aggregation and post-processing of retrieved information from different sources readily possible. This capability may come into conflict with legal constraints and is bound to affect the traditional marketing of information. It will lead to the extraction of higher forms of intelligence from text and data. The user community is seen to shift from expert information specialists, who act now as middlemen, to the end-users of information. This less experienced user community will challenge the ingenuity of system designers for self-guiding, adaptive, and yet more sophisticated man-machine interfaces. The merging of wide-band digital communication networks with computer technologies will make it possible to interconnect computers, information centers, word processors, and other peripherals, worldwide. Techniques of tabular and graphical fact retrieval are examined. The prospects of fact retrieval by voice, touch screens, and videotext are discussed. The potential of two unusual three-dimensional display techniques, the computer-generated time-resolved integral hologram and the projection of virtual data images into space, are discussed. We conclude by examining the resulting problems and some solutions by example of our experience with the integrated Technology Information System at the Lawrence Livermore National Laboratory.

"Fact retrieval is the identification and use of information about events and measurements by techniques that increase our knowledge, understanding, and ability to simulate and predict social and natural phenomena."

1. INTRODUCTION

The 1980s may well be called the decade of information. Quick access to factual information, the ultimate product of the post-industrial society, has been given a boost in the 1970s by the merging of computer technology with communications.¹ This made it possible for us to generate, validate, and disseminate information faster and cheaper, and has brought it within the reach of any telephone.



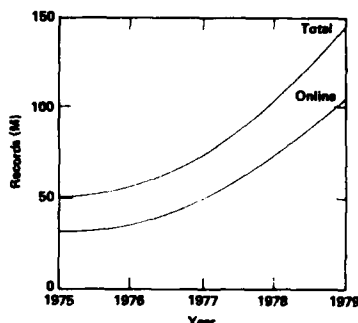
Traditionally, reports and books are used to convey the significance of an event or measurement. The embedded facts are then extracted and compiled in topical lists which serve as concentrated knowledge databases for comparison, evaluation, dissemination, and as a starting point for future work. These compilations of facts are structured to retain the essential attributes of their origin and qualify each individual fact for later use without descriptive text. It should not surprise us to observe that the historic use of computers for storage and retrieval of general information in science and technology has followed the same course.

Textual and bibliographic databases were created first. This marked the beginning of widespread, automated text retrieval services by machine, although digital computers are innately better suited to process numbers. The large bibliographic collections, now fully up-to-date in most fields of interest, represent a comprehensive online index to the books and reports of our civilization. By the end of the 1970s, 1500 discipline-oriented files provided quick reference to 150 million citations of the abstracted literature. Of these, some 450 databases were identified as being available online. They are the "sine qua non" foundation on which we can build the evaluated, numeric data files. Indeed, the relative ease by which text could be edited interactively and printed in different fonts, gave rise to numerous computer-aided typesetting techniques, a great improvement over previous manual and mechanical means to set type, although printing and distribution of the literature continued by traditional means.

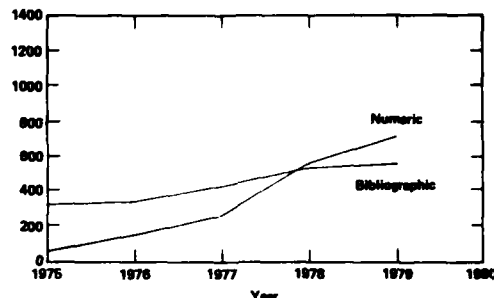
The creation of numeric data files of general interest followed. One of their categories is the huge collection of files created by automated sensors to describe the time- and space-dependent phenomena of demographic and environmental observations. Evaluated data of material properties, which are costly and complex to generate and to retain by computer in traditionally accepted forms, have made their appearance for use by a broader user community only in recent years. The total number of fact files has been estimated to exceed 10,000.

¹ Work performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under contract number W-7405-ENG-48.

INFORMATION CONTENT OF DATABASES:



NUMBER OF AVAILABLE DISTINCT DATABASES:



This trend toward numeric data files is expected to receive competition from electronic word processors. The rapid generation of written human communications by machine, and their sharing over wide-band, digital communications networks with other machines and powerful computers, are bound to create a natural mix of factual information, both textual and numeric. To work productively with these facts in large volumes will require the extraction of higher intelligence in concentrated forms. The problem is magnified by the fragmentation of the original information, dissimilar formats, and the necessity to present results in concert. This will challenge our ingenuity for innovative techniques of fact retrieval in the 1980s: How can we best establish practical procedures for the creation, storage, identification, validation, and display of the diversity and masses of factual data convincingly and with relative ease?

The immense power derivable from rapid access to accurate and up-to-date economic and technical information has been recognized by corporations and nations. Information management by machine in multinational business enterprises is reported to have increased corporate profits by as much as 30% in 1-2 years.² It is our hope that the apprehensions expressed at the recent 7th international CODATA conference in Kyoto, Japan, alluding to the potential prospects of data piracy and information monopolies in science and technology,^{3,4} may give way to a mutual and controlled sharing of this emerging powerful resource. This will help developing countries to catch up, increase the productivity of industrially developed nations, and benefit mankind, in general, rather than a privileged few.

In this paper, I would like to review the status of fact retrieval in science and technology as I have had the privilege to observe and practice it at a large national laboratory where advanced computers and communications are commonplace.⁵ In the parlance of Herman Kahn, this should make it possible to make surprise-free projections for the years to come.

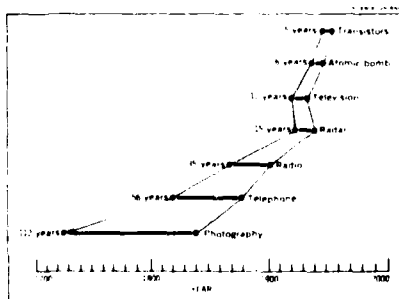
2. THE STATUS OF FACT RETRIEVAL TODAY

There is a saying that we must learn from the past, lest we be damned to repeat its mistakes. Our minds filter bits of information continuously. We compare them with past experience, establish their worth, and remember the essential facts for future use in an unending chain of analysis and synthesis. For these bits of information to be valuable, they must be factual, up-to-date, accurate, accessible, concise, usable, and controlled.

"Fact and Value"

Valuable facts are more than numbers alone. They can be descriptions of events, or well articulated and accredited postulates. For example, the letter written by Albert Einstein to President F. D. Roosevelt⁶ expressing apprehension that scientists in Germany may have split the atom, and that it should be possible to build a bomb because more neutrons were thought to be liberated in the fission process than were lost or absorbed, and that the U.S. could preempt them. This letter did not contain a number. But it was so valuable at the time that it started the enormous Manhattan Project, led to the early end of World War II, and ushered in the nuclear age. I am pointing this out to emphasize that fact retrieval need not be limited to numerical data to be of value. Indeed, the extraction of higher intelligence from descriptive text may well be of greater worth. It is capable of projecting our dreams and plans farther into the future than the knowledge of a more accurate event or measurement alone.

The extraction of new insights from large volumes of information, be they text or data, is clearly the challenge of the future. The real payoff in fact retrieval from the literature and numeric data files is, therefore, not only the speed by which this retrieval can be accomplished, but the new insight and understanding to be gained.



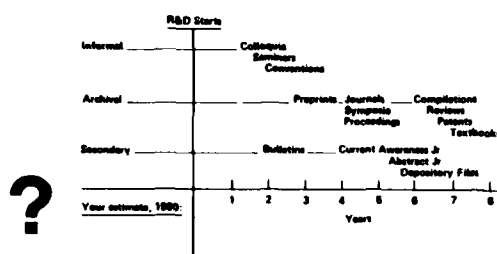
Shrinking time gap between initial discovery and final development.

"Datedness and Accuracy of Facts"

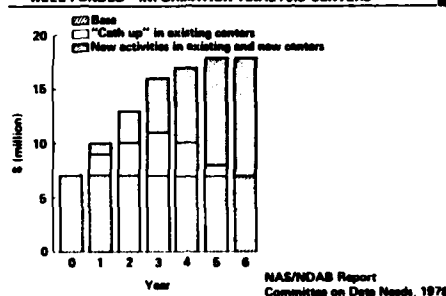
Our judgement is only as good as our knowledge! And yet, our information systems today are still hampered by long delays before factual information enters the main stream of computer-aided retrieval and evaluation. Estimates made in a report by the American Physical Society range from 2 years, for announcements of ongoing R&D in Bulletins, to 7 years, for compilations of evaluated numerical data derived from measurements. These estimates were made in 1970 but delays are probably not much shorter today. In the United States, little incentive has been given by government or industry to correct this situation. Other countries, more dependent on the flow of know-how and factual data from abroad than the United States, where 57% of all publication in the energy field originated in 1980, persisted in putting into action deliberate plans for national information systems.^{7,8,9,10}

UP-TO-DATE INFORMATION IS DIFFICULT TO COME BY

American Institute of Physics, 1970:



ACCURATE DATA REQUIRES WELL-FUNDED "INFORMATION ANALYSIS CENTERS"

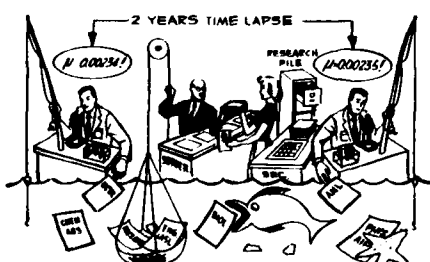


This inadvertent delay before results of ongoing research can be retrieved by computer is further exaggerated in the United States by the inadequate funding of Information Analysis Centers which compile and evaluate the measured data. Recognized data evaluation centers, such as those at the NBS Office of Standard Reference Data, and others operating as members of the Standard Reference Data System, have been working for years with limited and inadequate budgets. The Numerical Data Advisory Board of the National Academy of Sciences, Committee on Data Needs, published in 1978 a report on National Data Needs for Critically Evaluated Physical and Chemical Data (CODAN).¹¹ It recognized an urgent need for the doubling of support for data evaluation over a 6-year span of time just to catch up with the backlog of existing measured data. The overall national funding for data evaluation in 1977 was only \$6.8 million, of which the government provided 90%. Little action of any consequence has been taken to date. Other countries, however, are taking the initiative to capture the transborder information flow and are establishing in new and old areas of R&D their own information analysis centers. We are faced, therefore, with a situation where data are being measured with automated, computer-based equipment in larger quantities than ever before, but where the users of these data in the United States are obliged to fish them out from the general literature and make their own value judgements of good or bad, to remeasure them, or to buy data from abroad where available.*

"The Cost of Facts"

It is difficult to attach a price tag for facts derived from news or events. Their value is determined by timing and circumstance, as was pointed out earlier with reference to Einstein's historic letter.

Primary publications are priced by page regardless of their value:



SEA OF KNOWLEDGE

Page Range	Domestic Price	Page Range	Domestic Price
001-025	\$ 5.00	326-350	\$ 18.00
026-050	6.00	351-375	19.00
051-075	7.00	376-400	20.00
076-100	8.00	401-425	21.00
101-125	9.00	426-450	22.00
126-150	10.00	451-475	23.00
151-175	11.00	476-500	24.00
176-200	12.00	501-525	25.00
201-225	13.00	526-550	26.00
226-250	14.00	551-575	27.00
251-275	15.00	576-600	28.00
276-300	16.00	601-up ¹	
301-325	17.00		

Add 2.00 for each additional 25 page increment from 601 pages up.

Online bibliographic references that point to the fact that potentially relevant publications took place are sold with even less discrimination at constant unit price, regardless of value or length of their primary publication:

Offline printing	\$0.05	—	\$5.00 per citation
Online printing or viewing	\$0.50	—	\$50.00 per citation

Online numeric data are not sold on a unit basis today. As a rule, they are part of a larger file and costs are based on algorithms of Computer Resource Units, storage costs, and connect times, as will be shown later on. As such, they are much cheaper today than the primary publication wherein they are embedded, or the online citations that point to them, even though these facts are the essential and costly part of any document.

* Col. A. Aines, longtime advocate of data management in science and technology by computer has extended a challenge, to any one who could, to compile a report of horror stories where available data were not found in time, or were remeasured at great expense; or even better, where the wrong data were used and caused the failure or delay of costly projects. So far, no one has volunteered. Those involved are too embarrassed to speak up. Perhaps we should ask the retired members of our professional societies to speak to us with courage of their mistakes and lessons learned.

Therein lies somewhat of a paradox: Demographic and time-series data of our environment are massive, and their value may become quickly obsolescent. But, when we look at the cost of accurately measuring material properties in science and technology, we arrive at a very high unit cost. In the previously mentioned CODAN report, a literary publication in the sciences was estimated to cost \$45,000 on the average in 1976 dollars and required \$4,000 for printing. This was estimated from the time and effort required to do the measurements, assuming a year's worth of research and a median annual salary of about \$22,500 with 100 percent overhead. The absolute worth of data in any publication is difficult to judge in the absence of compilations and comparison with other data by expert evaluators. To establish their worth obliges the potential user to familiarize himself with the subject matter, to review the literature, and to arrive at an unbiased value judgement. This points to the urgency and overall cost-effectiveness of authenticated data evaluations.

The cost for compiling and evaluating data is comparatively small. Based on 16 years of operating experience with the Joint Army Navy Air Force (JANAF) Thermochemical Tables, the unit cost of data, representing one material property as a function of temperature and/or pressure, is only \$1,000. Since ten publications are found to support one data sheet in most cases, the investment of \$1,000 can capture the essence of \$500,000 in R&D expenditures for others to use with confidence.

The scarcity of up-to-date evaluated data compilations of physical and chemical properties probably accounts for their low market value. It leads inadvertently also to an unaccounted transborder flow of costly primary data. Other countries are making it their business to harvest these factual resources and to market them with advantage!

The Storage of Facts*

Storage requirements of facts by electronic means are assuming staggering proportions. We are committing more written communications, observations, and calculated data to computer-aided storage and retrieval than ever before. Estimates for some of the large data producers and users are:¹²

Lawrence Livermore National Laboratory	50,000	Tapes
Bank of America	400,000	"
Shell Oil Development	600,000	"
Social Security Administration	750,000	"
Exxon Geophysics Logging	1,000,000	"
	2,800,000	"

This represents (at a storage density of 6250 bits per inch for a 0.5-in.-wide magnetic tape of 2400 feet, inclusive of a 30% waste due to record gaps), a conservative capacity per tape of 1×10^9 bits, or an overall requirement for the above sample of organizations of about 10^{15} bits. Only the IBM Photo-Digital storage systems with 10^{12} bits capacity each, came close to filling such a demand; but, all of these mechanical, wet-chemical systems are now retired. Today, as we are entering the 1980s, the projected state-of-the-art for storage technology is estimated to be:

Magnetic	10^9	bits/cm ²
Optical	10^{10}	bits/cm ²
Electronic	10^{11}	bits/cm ²

Of these, only the magnetic tape storage media represented by Automatic Tape Libraries (ATL) are marketed. To appraise the situation at the DOE National Laboratories, and to make demand forecasts for future storage requirements, a survey was made in 1979 with the following results:¹³

Survey Results on Storage Capacity at Nine DOE Sites for 1979.

Site	Tape		Maximum tape capacity (10^{12} bits)	Disk drives	Disk packs	Maximum disk capacity (10^{10} bits)	Mass storage	
	drives	Tapes					archive storage	used (10^{12} bits)
Argonne	24	17,120	8.07	128	396	24.0	none	-
Brookhaven	22	8,000	1.72	28	56	7.17	none	-
Idaho Falls								
CYBER System	8	18,000	4.68	12	20	2.76	mag tape	8 week
IBM 360/75 System	9	23,600	4.60	26	40	3.2		backup only
Fermi Lab	24	30,000	5.85	20	20	3.2	ATL	
LASL	44	50,000	6.5	70	75	6.0	IBM 1360	2.2
(fixed head)					53	12.7	IBM 3850	2.3
LBL	22	41,500	7.07	24	24	2.82	IBM 1360	3.2
		+ 30,000 archived					IBM 3850	
LLL	50	45,000	4.41	110	175	152.0	IBM 1360	3.3
							ATL + CDC 38500	1.0
Sandia/ALB	45	26,000	1.74	77	150	1.51	none	
Oak Ridge	39	36,500	6.54	134	430	39.0	none	
Total	286	125,720	51.2	629	1009	754.4		

* A data sheet may be a single reaction rate constant as a function of temperature. By 1978, 2239 data sheets had been produced in 16 years.

Survey Results on Projected Storage Capacity at Nine DOE Sites for 1984.

Site	Type drives	Tapes	Maximum tape capacity (10 ¹² bits)	Disk packs	Maximum disk capacity (10 ¹⁰ bits)	Mass archive storage	Expected capacity (10 ¹² bits)
Argonne	24	22,120	11.6	104 ^a	15.7	SOA	0.32
Brookhaven	30	8,000	8.06	50	20	SOA	5
Idaho Falls							
CYBER System	14	28,000	8.84	42	72	SOA	...
IBM 360/75 System	13	33,000	10.8	9	4.3	SOA	...
Fermi Lab	40	55,000	55.4	60	9.6	ATL	9.07
LASL	50	50,000	50.4	80	46	tape cartridge	6.5
LBL	14	20,000	11.4	14	2.4	optical disk	2
LLL	17	10,000	10.1	16	35.8	optical disk	100
Sandia/ALB	32	18,000	11.4	120	1.2	ATL	1
Oak Ridge	2	1,000	0.15	4	1.3	tape cartridge	0.4
Total	236	245,120	178.2	499	208.3

We observe an overall increase of anticipated archival storage by a factor of ten for 1984, and a simultaneous decrease of installed equipment. One of the interpretations offered is an expectation that optical devices should indeed be capable of providing 10 times the storage density now possible on magnetic tape. The results of this survey were used to set specifications for storage devices in the mid 1980s. However, to date we have received no indication from serious respondents. An extract from our specifications follows.

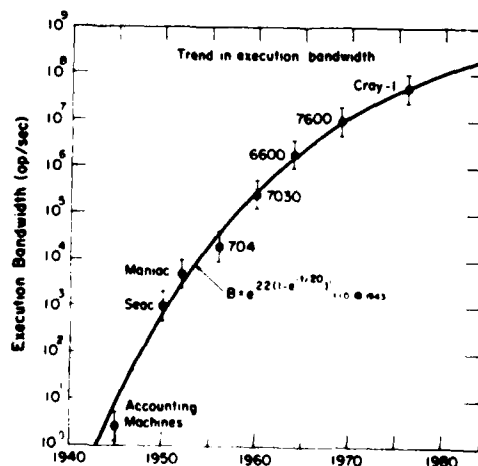
EXPECTATIONS FOR ARCHIVAL MEMORY PERFORMANCE¹⁴

Property	1980	1982	1984
System capability (bits)	$\geq 2 \times 10^{11}$	$\geq 10^{13}$	$\geq 10^{15}$
Read-write stations (No.)	≥ 2	≥ 4	≥ 10
Read-only stations (No.)	≥ 2	≥ 10	≥ 30
Accesses per hour	~ 2000	~ 8000	~ 30000
Transfer rate (bits/s)	$\sim 10^6$	$\sim 4 \times 10^7$	$\sim 10^8$
System bit error rate	$\leq 10^{-11}$	$\leq 10^{-12}$	$\leq 10^{-14}$
Undetected bit error rate	$\leq 10^{-12}$	$\leq 10^{-14}$	$\leq 10^{-16}$

The rates at which data are being added today are equally staggering: At LLNL, anyone of the four CRAY computers can generate 10^{10} bits of data per hour for a typical 2-D hydrodynamic calculation. As a rule, only 1/10 of that, or 10^9 bits, need probably be saved. A yet larger data source is the Landsat satellites where one pass would produce 10^{16} bits of data, were it not for selective data suppression and compression. But the techniques of file retrieval from these large storage media are not our concern in this paper. They are specialized and are not likely to affect the general user requirements of fact retrieval. Most of our needs in the 1980s will probably be well satisfied by extraction of facts from prestaged archival data files, and from interactions with less powerful, but more numerous minicomputers that have entered our working environment.

"The Analysis of Facts"

Programmatic requirements of the Department of Energy have contributed significantly to the design and development of high-speed computers. Each new generation of the fastest machines has been used for data analysis and to simulate intricate models of nuclear weapons, laser fusion, magnetic fusion, reactor safety, biological/environmental phenomena, and socioeconomic energy predictions. But the rate of performance improvement is slowing down. It took all of the 1970s to gain another factor of ten.¹⁵



Recent advances in Josephson junctions and the Gaits technology suggest yet another quantum jump to 10^9 operations per second in the 1980s. The significance of this expectation is that another generation of even more powerful micro- and minicomputers may also become available for general use. This would accelerate the dramatic changes of the traditional marketing and utilization of bibliographic/numeric data that we are seeing today. The post-processing of retrieved information toward higher forms of intelligence is clearly the challenge of the 1980s.

"Directories to Databases of Facts"

To access and analyze information, one must first know where to find it! The publishing business traditionally prints periodic updates of reference books on topical issues. In recent years, these publications were augmented by printed directories to computer-based resources. The current "Directory of On-Line Databases," by Cuadra Associates lists 770 data files and 135 on-line services.¹⁶ The "Directory to Computer-Readable Databases," updated each year by Martha Williams under auspices of the American Society of Information Science, shows in its 1979 edition 528 distinct data files.¹⁷ The corresponding international directory to computer-based systems with European emphasis, EUSIDIC, published by Learned Information, Inc., lists more than 1280 databases.¹⁸ The somewhat dated report by Westbrook on data sources for materials, cites in 1978 about 300 data files in science and engineering.¹⁹ Unfortunately, these directories are not yet available online even though computers were used to compile them and to print them. This is bound to change in the near future as government and business recognize the value of these master guides to information stores.

"The Tools of Fact Retrieval"

Data storage and retrieval was the goal of the 1970s, and different database manipulation software was built commercially and at universities to bring it about. We distinguish two classes:

- 1) Data Management Systems (DMS), which permit access to and retrieval from already existing files, usually for single applications. Most bibliographic online information retrieval systems are of this type.
- 2) Database Management Systems (DBMS), which manage and maintain data in a prescribed structure for the purpose of being processed by multiple applications, independent of storage device class or access method. They organize data elements in some predefined arrangement in a database and retain relationships between different data elements within the database. These systems are commonly used with numeric/structured data under the user's control.

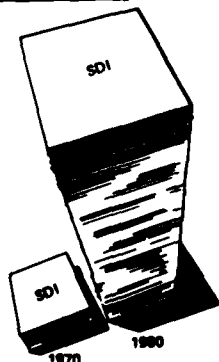
For large volumes of data, like those being communicated to earth from observation satellites, efficient and specialized programs were developed and are usually not suited for generalized applications by casual users. For comparatively small and diversified collections of data, a host of more generalized, less efficient data management systems have evolved. Depending on the logical relationship among data sets and data elements, the major DBMS models now in use are those best capable of working with:

- o simple, COBOL-like 'flat files',
- o hierarchical or tree-like data,
- o graphs and networks, (CODASYL),
- o relational tables, and
- o data well represented by extended set theory models.

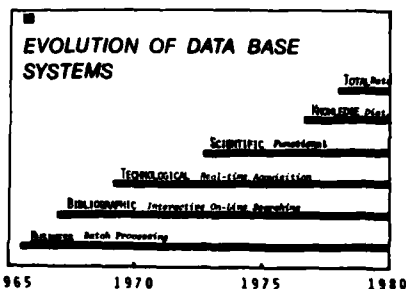
No definite bounds exist among the systems that implement these models. Most systems show considerable flexibility. No consensus has been reached to define the best model for general database work. An overview of their relative merits and historic dependence was made by Fry.²⁰ But, to those shopping for potential, suitable candidates among the large number of systems, their underlying theory, strengths, and limitations may only be of partial interest. The questions more likely to be asked concern user-oriented features; the types of computers and operating systems on which the systems can be installed; the host languages to which they interface; and cost.

The online bibliographic retrieval systems and services have changed little since their inception in the late 1960s. Only the numbers of online topical data files and citations have quadrupled. Many still seem to believe that the greater the number of citations retrieved in retrospective searching, or by Selective Dissemination of Information (SDI), the better the service. Relevancy, recall, and precision were hotly debated issues and vary with the sophistication of the indexing and retrieval method, and whether keywords, and/or titles, and/or abstracts are scanned. At LLNL we tried and documented several experiments and settled on searching by indexed keywords and words in titles. Abstracts introduced too much noise. Most of the bibliographic systems are patterned after DIALOG (Lockheed) and provide Boolean operators for retrieval of whole words or compound expressions from indexed tables of keywords and authors. Sets can be created and combined to further refine a desired result. Few bibliographic information management systems offer capabilities for interactive algebraic work with numbers. Free-text searching of titles or abstracts is less frequently available. But, regardless of how these systems retrieve their citations, nearly all bibliographic information today is being delivered as a pile of paper. At best, it is printed or flashed on the CRT screen in reverse chronological order of its publication date, requiring the hapless recipient to just look at it!

Traditional SDI Service:



"Look, how much I found for you..."
 "You can't store it in your mini..."
 "You can't reproduce it by any means..."
 ... just look at it..."



Retrieval with embedded character strings, weighting of terms, and proximity indicators have become more prevalent with the large legal text retrieval systems, e.g. JURIS, LEXIS, and WESTLAW, among others.

A detailed comparative analysis of numeric/structured data management by means of the commercial systems, or by university systems, is difficult because of the continuous change exhibited by these systems, estimated to number about 65 worldwide.²¹ Suffice it to say that most of them tend to be modular with interfaces to procedural languages and to statistical and graphical programs. Excellent seminars on "Comparative Database Management" are being offered each year by the University of California Extension Division, Los Angeles. Comparisons of 21 systems and their capabilities were made by Auerbach²² and Datapro.²³

One of the most promising retrieval systems for numeric/structured data developed during the 1970s for general use is the Chemical Information System, (CIS), sponsored by the National Institute of Health (NIH) and by the Environmental Protection Agency (EPA).

CIS is a network of chemical databases equipped with computer programs that permit interactive searching and retrieval from these databases. At the heart of the network is the Structure and Nomenclature Search System (SANSS) which is used to identify a chemical substance, given its name or its structure, and refer the user to all CIS files that contain data on the compound. Any such data can then be retrieved with simple commands from the appropriate node of the CIS network. A different way of using the CIS is to enter experimental data, such as mass spectral peaks, into the system, which will identify the compound at hand, using its files of ten of thousands of mass spectra, nmr spectra, or x-ray powder diffraction patterns. The entire CIS has been developed by cooperating agencies of the U.S. Government and is available in the private sector for use by the public on a fee-for-service basis. It is accessible worldwide through the TELENET telecommunications network. Today, the CIS system is still a collage of files and programs originally provided by developers of each database. However, gradually, a unifying approach is being implemented that will make CIS the most comprehensive collection of physical and chemical data files.

The CODASYL reports,^{24,25} published early in the 1970s, identified common and desirable features of database management systems and had a very beneficial impact on their evolution. They provided a basis for the database architecture and Federal database standards.²⁶ Generalized Database Management systems (GDBMS), tailored to account for attributes and peculiarities of data in science and technology, did not materialize as yet. In most cases they were adaptations of systems used in business. Attempts to point out the special requirements for scientific data were made by the specialists' conference on "Generalized Data Management System and Scientific Information" in 1978, sponsored jointly by the Department of Energy Office of Technical Information and the Agence de l'OECD Pour l'Energie Nuclaire.²⁷ This was followed by a NASA-sponsored conference on Engineering and Scientific Data Management²⁸.

For those who may wish to further explore the evolution of fact retrieval during the 1970s, I would like to call your attention to the success of the automated indexing of the bibliographic citations by the Defense Documentation Center,^{29,30} the experiments of data "tagging and flagging" still ongoing by the American Institute of Physics,³¹ and early attempts at integrating text and numeric data.³² Additional topical literature is found in the Annual Reviews of Information Science,³³ especially in Volume 14 of the 1979 proceedings of the "Online Information" meetings,³⁴ and the international conferences on very large databases.³⁵

"The Marketing of Facts"

Fact retrieval from commercial vendors costs about twice that from government installations where recovery of expenditures became an operational requirement during the 1970s. In the last few years, the pricing has remained reasonably stable, except for costs of connect-times. We quote from the Cuadra Associates Spring 1981 Edition of the online catalog:¹⁶

"Pricing policies for access to and use of the online database services are extremely complex. There are a number of components to the prices and they are combined in many different ways. In addition, prices are subject to change with fairly short notice. All of these factors make it difficult to treat the topic in a standard manner. However, there are some general points that can be made. In at least 30 percent of online services, there is an indication that some type of subscription is required for gaining access to the database. These subscriptions range from a few hundred dollars per year to several thousands of dollars. In some cases, the user subscribes to a package, which may include one or more databases and additional services (e.g., consulting). In other cases, the user has several options, each a combination of a subscription price and an associated usage charge.

In general, the major components of the usage prices for online database services differ according to the type of supplier and the type of database (e.g., whether it is a bibliographic database or a numeric database). There are two major groups of policies: one for the timesharing firms and their (largely) numeric databases, and one for the others, covering most of the other types of databases. There are, however, exceptions in each group, e.g., where a timesharing firm offers service on a referral database but prices it more like a bibliographic database.

Pricing by timesharing firms in business primarily for numeric databases, requires for most a monthly minimum (e.g., \$100 per month) that is applied if the total usage charges for a given month do not reach the minimum level. The usage charges include the following components:

Connect Time	\$1.00 -	\$21.00	per hour
Computer Resource Units	\$10.00 -	\$1.25	per unit (varies with definition)
Disk Storage			varies

These rates can also vary within a specific service, depending on the speed of the terminal being used (e.g., 300 or 1200 baud) and the time of day in which the processing occurs (e.g., prime time vs non-prime time). In addition, the Computer Resource Units charged for a particular database may be greater than the standard timesharing rates charged for other data processing services. This difference occurs either because the database system that is being used is more demanding of resources, or because a surcharge has been added to the standard rates (by a multiplier or additive factor) as a royalty to the producer of the database. In most other cases, pricing by online services is based on an hourly connect-time rate, plus telecommunications costs for network access, if applicable. The hourly

connect-time rates that are cited in the supplier's literature may include the royalty, or the royalty may be cited separately. The range of connect-time rates (including applicable royalties) is from about \$25 to \$300 per hour. The average is approximately \$65 per hour.

For bibliographic and some referral databases, there is an additional charge for offline printing, which is generally based on the number of citations or pages. The range of charges for offline printing is from \$0.05 to \$5.00 per citation, although for a few databases they may be considerably higher. The average is about \$0.15. There may also be charges for online printing (i.e., displaying retrieved information directly at the terminal). These fees range from \$0.50 to \$50.00 per item. Both online and offline printing charges may vary depending on the amount of information that is printed. In some cases, the use of a service involves a startup fee, which often covers account setup, initial training, and materials. Occasionally this startup fee also includes the cost of a special terminal and/or leased lines to the online service's computer. Many of the online services that focus on the provision of database access also provide volume discounts or have subscription plans that provide for various levels of connect-time rates, depending upon the expected level of usage."

The cost of using the CIS is approximately \$45 per connect hour for most components, \$75 per connect hour for the major files, notably SANSS and PDSM. In addition to the search costs, there is an annual \$300 subscription fee, which is used to defray all storage costs. Costs now levied by other government information centers are similar.

Search times per database require on the average about 15 minutes and are reported to range from just a few minutes to more than an hour. If we consider the 5 million searches, or queries, conducted in 1979 against the 150 million records of computer-readable files in the United States and Canada alone,³⁶ one arrives at an estimate for the total commercial revenue of marketing information on-line: \$125,000,000/year. Actual costs to the buyer who provides the information specialists, terminals, and research facilities, are much greater.

"Other Issues of Fact Retrieval"

There are other issues which have their origin in the 1970s: Wide-band communications, the electronic office, user-friendly interfaces, security, copyrights, and novel ways of fact analysis and display. Because of their significance in the years to come, I treat them jointly by discussing their expected state of technology in the next decade.

3. PROSPECTS FOR THE FUTURE

The rapidly increasing high-speed digital communications are starting to link previously separate communication channels. Voice, data, and video are beginning to serve as an integrated medium for fact retrieval. Their analysis and synthesis on powerful micro- and minicomputers will bring about new forms of communicable intelligence. This, in turn, will increase the value of factual information and introduce controls for its exchange and use.

"The Communication of Facts"

Faster and cheaper computers forced increasing demands for high-speed transmission of digital data in the past decade. Rates increased from 9,600 bps for early telephone lines to 12-14 GHz for satellites approved in January, 1981, by the Federal Communications Commission (FCC) for Satellite Business Systems (SBS). The impact of this change in a relatively short span of time will even be greater when the additional 20 domestic satellites authorized by the FCC last December come into operation. This will permit full integration of voice, data, video, and image transmission. The advantages are bound to affect every-day communication for fact retrieval:

- 1) Communication costs will cease to be a function of distance.
- 2) Point-to-multipoint broadcasting will make simultaneous updating of distributed databases practical.
- 3) Organizational, computer-based networks could be established virtually overnight.
- 4) The higher frequencies will permit the use of smaller earth-station antennas, 5-7 m in diameter.



At the Lawrence Livermore National Laboratory we now use two antennas with the WESTAR satellites which link the Magnetic Fusion Energy Computer Center (MFECC)³⁷ network with Princeton University. However, for most of us, it will take several years before the required number of ground stations are up and ready for use, and before the bandwidth for linking to these ground stations can be increased by local hyperchannel networks or fiber-optics communications, among others.

Today, as this paper is going to press, costs per month for cross-country communication between San Francisco and Washington, D.C., are typically those shown below:⁺

RATE	ANALOG		DIGITAL
	With Modems	Without	
1,200 bps	\$ 2,140	\$ 2,048	Not Offered
2,400 bps	2,218	2,048	\$ 2,124
4,800 bps	2,430	2,048	2,338
9,600 bps	2,756	2,048	2,684
19,200 bps	5,942	2,048	Not Offered
56,000 bps	25,499	2,048	11,286

* Vendor sources unknown, conditioning would be required.

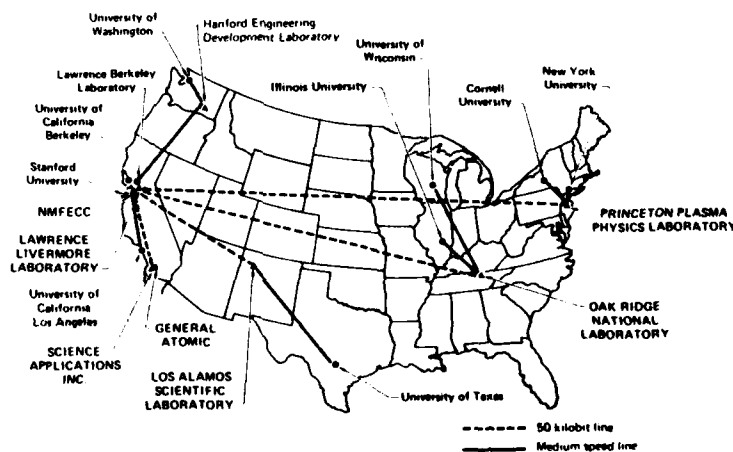
** No source available for vendor-supplied Data Service Units (DSUs) required to operate on the D.D.S. network.

+ Private communications with Jack Hibbard, LLNL, Communications Office, July 1981.

The rates quoted are with and without Bell modems. Installation charges for analog communication are \$100 without modems, and range from \$254 to \$639 with modems.

The digital computer networks ARPANET, TYMNET, and TELENET, made their entry in the 1970s. Of these, ARPANET is sponsored by the Department of Defense and was the first large-scale experiment in computer communications by linking various large machines on many university campuses and government installations. Although successful for medium-size file transfer between computers, the 50K-bit lines and packet switching techniques are inefficient for low-speed terminals operating in full duplex. This does not really matter since traffic on the ARPANET is still comparatively light. (Total number of bits moved on the MFECC net exceeded that of the ARPANET already in 1974.³⁸ Nevertheless, ARPANET became one of the most widely studied and publicized networks.

NATIONAL MFE NETWORK 1981



TYMNET developed quite differently. Its primary purpose was to interface large numbers of low-speed terminals to a relatively small number of time-shared computers operated by Tymshare, Inc. Most of these terminals also required full-duplex, character-by-character interaction with the machine. The inefficiencies echoing characters back and forth were solved by virtual circuits where data from many users shared the same physical record during transmission, so that the overhead of checksums and record headers could be divided among the users. Also, by controlling flow node-to-node, rather than circuit-end-to-circuit-end, there was no need to signal back a message requesting more data. This efficient operation permitted 40 300-baud interactive terminals to be served simultaneously by one 2,400 bps line. In 1972 the National Library of Medicine put the first non-Tymshare computer on the network. Since then, this Value Added Network (VAN), providing automated speed, code, and protocol conversion among terminals and computers, and error recovery by routing records around inoperative paths, has become one of the largest networks with more than 25,000 daily users in 1979, and connections to overseas.

TELENET commercialized the ARPANET concept in 1975 and was approved as an international record carrier. It interconnects to TYMNET and TRANSPAC in Europe, and the Canadian CNCP and TCTS networks. These three networks will continue to have a significant role for ground communication, and interconnect to satellite traffic as required. Existing and planned cable-video connections are expected to assume a significant share of the digital communications market.

The interconnection of these and other national carriers has brought about an international worldwide network. Hilsenrath at the National Bureau of Standards has recently surveyed the field and reports 250 different numerical data systems of which 53 deal with physical and chemical properties.³⁹ The networks are TYMENT, GTE TELENET, CYBERNET, GE GEISCO, and CYPHERNET.

"Standards for Data Exchange"

The opportunity to communicate computer-based information in large quantities introduced the necessity of standards.

For communications among computers, the X.25 communications protocol is probably one of the most remarkable achievements of the past decade. It has simplified the physical interconnection of host computers over networks. However, the interprocess communications among computers requires either a translation of often incompatible primitive operators, or their standardization. Both approaches are presently in progress.^{40,41} The Systems and Network Architecture Division of the National Bureau of Standards has the mandate to study the issues involved and to formulate standards in digital communications. The related problems and opportunities for the then Energy Research and Development Administration (ERDA) were studied and reported by the Working Group on Computer Networking of the Office of Engineering, Mathematics, and Geosciences.⁴² Problems and solutions of distributed data management and computer networks are reviewed each year in conferences at the Lawrence Berkeley Laboratory.⁴³

For communication of facts among dissimilar databases, several standards evolved in the 1970s:

ANSI Z39.2 - 1971 American National Standard for Bibliographic Information Interchange on Magnetic Tape.

ISO 2709 - 1973 Documentation - Format for Bibliographic Information Exchange on Magnetic Tape.

ISO 646 - 1973 7-Bit Coded Character Set for Information Processing Interchange.

ISO 962 - 1974 Information Processing - Implementation of the 7-Bit coded Character Set and Its 7-Bit and 8-Bit Extensions on 9-Track 12.7 mm (0.5 in) Magnetic Tape.

ISO 2022 - 1973 Code Extension Techniques for Use with the ISO 7-Bit Coded Character Set.

ISO 2375 - 1974 Data Processing - Procedure for Registration of Escape Sequences.

ISO/DIS 1001.2 - 197x Information Processing - Magnetic Tape Labelling and File Structure for Information Interchange.

UNISIST SC74/WS/20 Reference Manual for Machine-Readable Bibliographic Description.

IAEA/INIS-9 (Rev 1) INIS: Magnetic Tape Specifications and Record Format.

TID-4581-R3 ERDA Energy Information Data Base: Magnetic Tape Description.

These standards deal primarily with bibliographic information. The growing importance of numeric data in the mid 1970s led to the "ERDA Interlaboratory Working Group for Data Exchange," which studied the characteristics of the transmission of numeric data. Based on the previously established standards for bibliographic information, the group proposed a numeric data exchange format, officially referred to as the proposal for an "American National Standard Specification for an Information Interchange Data Description File Format." In 1978, this draft became the substance for the ANSI X3L5 committee, which included several refinements and extensions recommended by international reviewers. The CODATA "Task Group for Computer Use" endorsed the standard and asked UNESCO to have it disseminated in other countries for potential acceptance as an international ISO standard. Formal transmission of the final recommendation by the X3L5 committee for ANSI confirmation is scheduled for July 1981. It may take an additional 6-9 months before the standard could be officially accepted, printed, and distributed.

"The International System of Units (SI) for Numeric Facts"

Shortly after World War II, in 1954, it appeared as if the United States might have the resolve to change from its English system of units of measurements to the metric rationalized and coherent system of units based on the four MKSA (meter, kilogram, second, ampere) units, plus the degree Kelvin as the unit of temperature, and the candela as the unit of luminosity. Although the United States participated in these international deliberations, extension of MKSA and the adoption of the International System of Units (SI) did not receive final approval before 1976.⁴⁴ Public and industrial support came even later. The advantages of only one unit for each physical quantity, a well-defined set of unique abbreviations and symbols, and the retention of decimal multiples and submultiples of the base unit for each physical quantity, have made it now possible to exchange numeric databases with less difficulty.

The new SI system of measurement is being adopted throughout the world. Its details are published and controlled by an international treaty organization. However, for a number of reasons, it is inevitable that a few other units outside the system be used with it. It is this additional use of non-SI units that leads to controversy and difference between standards that define modern metric practice. Since a wide variety of metric units have been in use for years in various parts of the world, it is natural that tradition would promote use of these old units in formerly metric countries. For this reason, many European and international standards also recognize a number of non-SI units for use. To protect the new system from degradation, and to cooperate with people all over the world, the SI standard is recommended for all work with numeric factual data.

During the transition period, where technologists are still accustomed to recognize the validity of material properties by their remembered values in former units of measurement, I believe that database management systems for numeric data should offer the option of displaying values in original units of measurement, SI units, or both. It is in this manner that our eyes and minds can be trained to learn, remember, and work with the new SI units with growing confidence.

"Standards for Fact Attributes"

Accreditation of facts, especially of numerical data and measurements, requires a "shorthand" notation with attributes which a scientist or technologist would accept in place of the original descriptive text. The minimum necessary and sufficient set of such attributes has been discussed in the literature. The table below is a representative list.⁴⁵

Attributes of Scientific Data

- 1 Value
- 2 Uncertainty
- 3 Units of Measurement
- 4 Normalization
- 5 Validity Domain
- 6 Method of Measurement
- 7 Conditions and Constraints
- 8 Type of Data
- 9 Source of Data
- 10 Bibliographic Reference
- 11 Comments
- 12 Proprietary Status, Classification, etc

One of the reasons why numerical databases in the pure sciences may not have come into greater use is precisely the lack of agreement on this topic. This may also explain why a recent survey carried out by the World Federation of Engineering Organizations still lists evaluated numerical property data banks as a desirable, yet less frequently used resource.⁴⁶ A number of relatively low-level efforts are in progress to correct this.^{47,48} A more authenticated approach is needed. David Lide, Director of NBS/OSRD, emphasizes in a recent publication in Science that standards are urgently needed for critical data and data banks.⁴⁹ Use of such standards for the reporting of numeric values for a single datum, and/or a set of data, in the different disciplines, could serve to establish confidence for both data evaluators and users. The NBS Computer Institute for Science and Technology and the Office of Standard Reference Data should document their recommendation in the existing series of Federal Information Processing Standards (FIPS).

"The Copyrighting of Facts"

Computer technology confounded the interpretation of traditional copyright laws. Unlike conventional printing techniques, computer-based systems were capable of storing, processing, retrieving, transferring, displaying, and reproducing works of authorship with ease. In addition, any one of the above processes requires, as a rule, the fixation of more than one copy, (core memory, disk storage, CRT display, etc.) and could thus be construed as a copyright violation. Also, authors and owners of copyrights of works argued that their creations should be patentable, because it is the idea, procedure, process, system, or method of operation that should be protected, rather than the form (i.e., source program, code) in which it is described, explained, or embodied. Section 117 of Public Law 94-553 of October 1976, therefore, did not confront these difficult issues and stipulated that computers did not afford the owner of copyright in a work any greater or lesser rights with regard to the use of the work by computers than those afforded under sections of the common copyright law. The National Commission on New Technological Uses of Copyrighted Works was established a few months later to study the matter. It issued the final report in July 1978, and recommended that Section 117 be removed and/or clarified to provide copyright protection for computer programs and databases.^{50,51} This was substantially enacted by Public Law 96-517 in December 1980, with the explicit provision that owners of a software product can copy, or authorize copying or adapting, the work without infringement on the copyright owner's rights if this action either constitutes an essential step in using the program with a machine, or serves archival purposes only.

The recommendations were based on public hearings with particular attention to the problems of copying computer-generated images, e.g., bibliographic citations. The problems of copyrighting numeric data were not explicitly addressed, and some confusion exists as to whether numeric time series data of demographic or environmental observations, and of measurements, can or should be copyrighted. Before I offer an opinion as a user of numeric data, I would like to make the following observation.

The rearrangement of the contents of a copyrighted book, and its printing and marketing in modified form in competition with the original book, would probably be an infringement of fair use, as defined in Section 107 of PL 94-553. A similar situation can be inferred for numeric factual data. Since any information or data with a particular embodiment, e.g., the attributes of uncertainty, precision, normalization, operating conditions, material history, etc., can be protected by copyright, their rearrangement and reproduction in whole or substantial part would probably also be interpreted as an infringement when it affects detrimentally the rewards derived or expected by the owner from the original copyright. This observation, if substantiated, would have significant implications on the three processes by which factual databases in science and technology are created: collection (aggregation), derivation (analysis), and compilation (evaluation). In each case, the authors and owners of these embodiments can request and receive copyright protection for their new works. Royalty payments to the owners of copyrights for the contributing works would have to be resolved in court, if not previously negotiated.

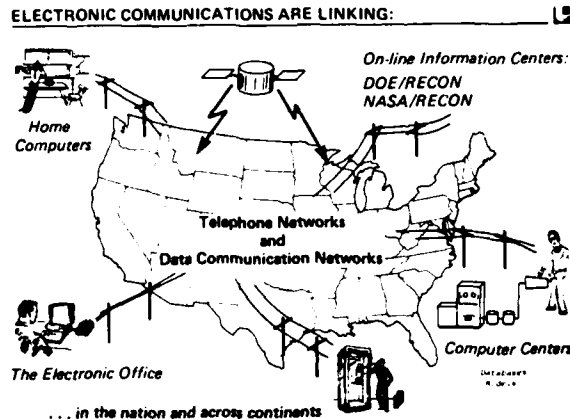
Let us look at some examples. The Department of Commerce makes available magnetic tapes containing numeric data of physical and chemical properties, evaluated under auspices of the NBS Office of Standard Reference Data. These databases are being sold to recover costs, and are copyrighted on behalf of the United States, as mandated in Sections 5 and 6, respectively, of the Standard Reference Data Act, also known as PL 90-396 (1968). Of the approximately dozen magnetic tapes now available from the Department of Commerce, only half contain material properties. The majority of the numeric data, measured and reported by the enormous, tax-supported research and development program in the United States, are, to the best of my knowledge, unprotected. Yet, they form the foundation on which compilers and evaluators in the United States and abroad build their topical databases which then are copyrighted and marketed without royalty payments to the experimentalists, the publishers of the primary literature, or the United States Government. Compilations of numeric data are thus treated similar to compilations of citations in bibliographies. The difficulties in data exchange and transfer arise only after such numeric compilations come into being and are copyrighted in a manner that is not in the interest of the United States.

In view of the importance of factual data in the 1980s, the copyright issue of numeric data, and the issue of transborder flow, will undoubtedly be very carefully examined.

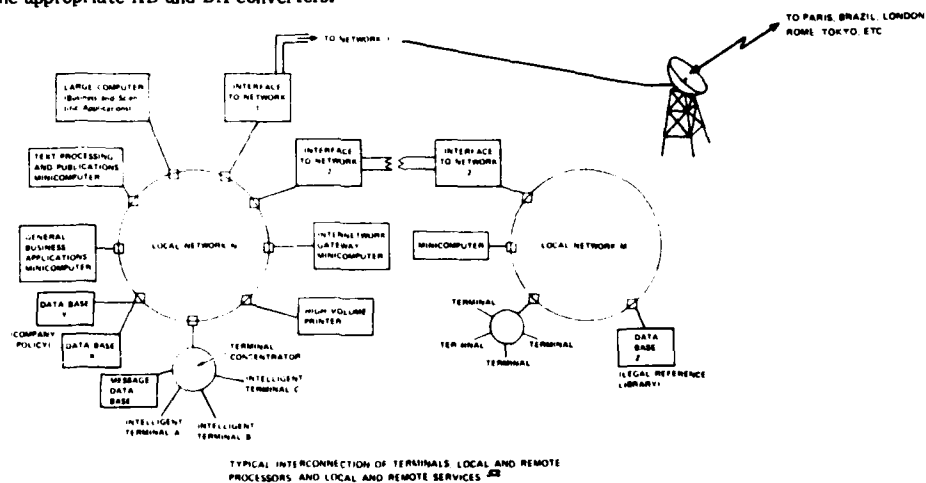
"The Electronic Office, Video Transmission, and Electronic Mail"

Much has been written on this topic.⁵² It is clearly big business when we are given the capability of interconnecting word processors with computers, typesetters, and graphics devices, and can send a resultant illustrated report cross country in seconds. Productivity and creativity are increased. Work satisfaction is enhanced also as secretaries and electronic word specialists now create camera-ready copy that matches the quality previously reserved for publishing houses.

At the present time there are two competing yet complementary telecommunication and computer-based technologies available for transmitting, storing, retrieving and disseminating large volumes of information electronically. One of these, the Videotex (Viewdata) concept, exemplified by such systems as Antiope, Prestel, and Telidon aims at providing general information to a mass consumer and general business audience and is in an early developmental stage. The other, exemplified by bibliographic and numeric information systems, has been discussed before.⁵³



Video image and video text transmissions require wide-band communications. These can now be realized locally, but will not be possible cross-country at reasonable cost before sharing of satellite communications becomes commonplace. (One 256K-bit picture requires at least 3.6 minutes for transmission at common 1200 bps speeds.) But, more than 50,000 pictures can be stored on a \$4,000 video tape or video disc machine when connected to an intelligent terminal and the appropriate AD and DA converters.



Electronic mail, usually understood as an expedient means of sending and receiving typed messages and reports, is going to be augmented by voice message systems. Voice input, although now available with limitations, will take longer to implement but is certainly going to become one of the modes of man-machine communication, together with touch-screen command selection as marketed, e.g., by Control Data Corporation in their PLATO system.

The conversion of video text into ASCII text by means of Optical Character Reader techniques is still to come. But, when it takes place -- and it is technically feasible now -- we will experience a total integration of:

Speech \longleftrightarrow ASCII Text \longleftrightarrow Video Text

The translation of printed text into speech is already being offered by the Kurzweil Reading Machine for the Blind.⁵⁴

"Encoding and Decoding of Facts"

Communications in the 1980s will require protection from eavesdropping and abuse. It is against the law to tap a telephone line. However, the upcoming swarm of communication satellites which is starting to blanket large geographic territories with data and video transmissions, virtually invites anyone with a parabolic roof antenna to "listen" in. User identifications and passwords, if transmitted in clear form, could be compromised. Governments, businesses, and banks have long ago learned to protect their interests by encoded transmissions. The flurry of sales of \$2K to \$15K roof antennas to harvest the multichannel free broadcasting of television programs will probably be short-lived.

Future users of satellite communications in science and technology will have the advantage of being able to choose from a variety of proven hardware/software combinations to protect their interests. I would like to draw your attention primarily to the public-key cryptosystems that have been proposed toward the end of the 1970s and are likely to have a significant role in the coming years. Their security emphasis has changed from statistical uncertainty to computational complexity. They have developed from conventional private-key cryptosystems to public-key cryptosystems, providing instant privacy and two-way authentication. Some of the essential observations, as summarized by Abraham Lempel, are:⁵⁵

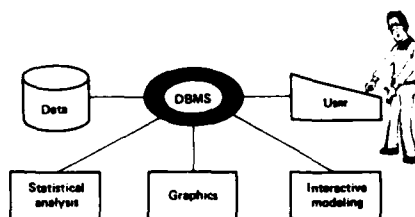
Cryptology and the computational power of electronic computers, once mainly concerned with the making and breaking of secure military and diplomatic communications, have now become a major concern of the public at large. In the age of ever-growing computer data banks and electronic fund transfer, one cannot overestimate the importance of encryption schemes which provide adequate protection against unauthorized, often remote, access to stored data, render the data unintelligible to unauthorized listeners over a publicly accessible communications link, and incorporate a digital signature which can serve as a reliable two-way authentication. These formidable goals were set up to satisfy real market demands, answered in part by the Data Encryption Standard (DES). This is the official NBS scheme to be used by Federal departments and agencies, as well as others, for the cryptographic protection of computer data. Those critical of the DES argued that computers in the early 1980s, rather than toward the end of the decade, would have the power to break the DES code in a day. Alternate schemes have, therefore, become more attractive.

The concept of public-key cryptosystems, introduced in 1976 by Diffie and Hellman, envisioned a system for private communication that employs a public directory in which each subscriber places a procedure E to be used by other subscribers for the encryption of their messages addressed to him, while keeping secret his corresponding decryption procedure D . The existence of such a system would enable instant secure communication between subscribers who have never met or communicated before. For example, if subscriber A wants to send a private message M to subscriber B, he looks up E_B in the directory under B, and transmits $C = E_B(M)$ in the open. Only B can decrypt C by applying his secret D_B to C .

One of the major shortcomings of currently practiced cryptography--the DES, as well as the new public schemes--is the lack of proof that any of these schemes are, indeed, as hard to break as they are claimed to be. Nevertheless, encoding and decoding of factual data will necessarily become a way of life in the 1980s.

"The User Interface"

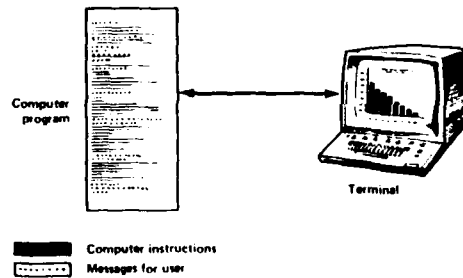
From the preceding review and projections, it becomes apparent that modern technology can link virtually any dissimilar pieces of hardware and software into integrated systems of higher purpose. Differences in format and standards cannot be avoided and do contribute to less efficient operation of the whole. But, they really do not pose a lasting hindrance. Economic pressures cause them to conform. The essential key required to unlock the enormous potential of stored information and factual data lies in the user interface. It is the mediator between man and machine. It makes the formidable aggregate of powerful computers and communications appear to be human, and it imbues us with qualities of exactness and precision more likely to be expected from a machine.



Different methodologies were employed in the past decade to accomplish this goal of translating English-like logical requests into machine instructions, and vice versa. In most cases, the interface has been tailored to serve well one particular Data Management System (DMS), Data Base Management System (DBMS), or some related analysis and graphics package. Although modular in nature, interfaces and their man-machine intercommunications usually became an integral part of the programs they served. Thus they were seldom capable of being extended to other systems, or flexible enough to accept the never-ending demands of an active user community. (DIALOG today is virtually the same as invented more than a decade ago.) References to one or the other approach can be found in publications and proceedings.

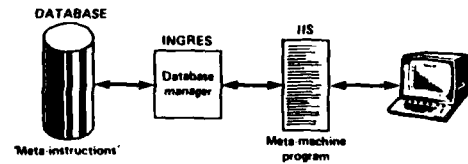
Here, I would like to report on the performance of a unique interface, the META-MACHINE,⁵⁶ designed and developed by LLNL in collaboration with Control Data Corporation. It serves as an extensible, flexible, and practical interface for the integrated Technology Information System described toward the end of this report. It translates pragmatic English commands into meta instructions for an open-ended number of programs which it controls. It can be readily adapted to languages other than English.

TYPICAL INTERACTIVE SYSTEM



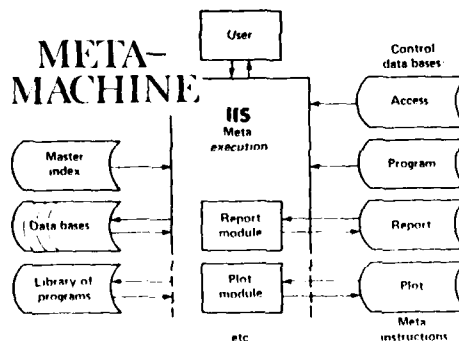
THE 'TIS' APPROACH

Messages and Computer Instructions Reside on Disk in a Database

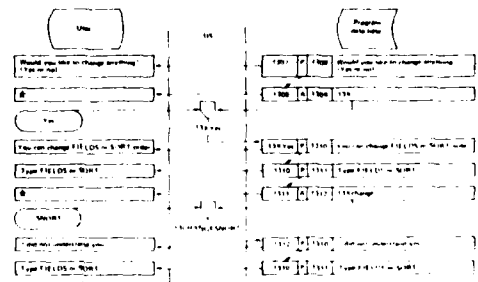


The META-MACHINE is fully self-guiding and supports interactive text and data retrieval, interactive modeling, electronic mail, and networking. It is the central controller for all man-machine and machine-machine communications. It is currently installed on a PDP-11/70 machine at LLNL, using INGRES as the relational database management system, and UNIX as the operating system (2.25 M bytes of core and 1 B bytes of disk storage).⁵⁷ However, INGRES and other subservient programs are invisible to the user. Unlike in other interfaces, all man-machine communications are deposited in an INGRES relational database and can therefore be changed in real time, online, without recompiling.

INTEGRATED INFORMATION SYSTEM



META MACHINE (SYMBOLIC EXAMPLE OF OPERATION)



The META-MACHINE simulates in software a pseudo-computer where sequential instructions are retrieved from a 5-domain program relation table. These domains, consisting of a primary address, a state function code, a forwarding address, an execution string, and a functional clustering attribute, can be arranged and manipulated by the System Administrator with standard INGRES database management commands. Meta instructions in the execution string contain prompting statements to be forwarded to a user, confirmations, command strings for the underlying database management system, or instructions for other programs.

The user is prompted by a uniform command language. His requests are matched against the available options in the hashed primary address domain of the TIS program relation, and are exchanged and assembled into corollary instruction strings for whatever program is to be executed. The transmittal of the prompting strings to the appropriate user at a terminal, or the transfer of an instruction stack to a program, is carried out by a 40,000-byte program residing in core. The 17 state functions determine the destination and type of string delivery. The following figures illustrate this functional procedure. During man-machine interaction with the user, while the command execution strings are being assembled, partial transmission to the target program, local or remote, can be made in anticipation of the full command string. In this manner, and by anticipating permissible attributes and parameters through the functional command cluster attribute, considerable speed can be gained.

The META-MACHINE approach uses one data management technique for access to all information, data, and executive capabilities. This includes control of access rights by user for each of the major databases, the individual relations within each database, the user commands, the preformatted reports and graphical displays, and especially the information about other centers to which it automatically connects.

The flexibility of the META-MACHINE has been demonstrated by the successful integration of bibliographic, numeric, project-oriented, administrative, and budgetary information or data files in one system. The report writer, exceeding the ANSI COBOL74 capabilities, is also driven by instructions from the META-MACHINE. Using this unified technique, the report writer was written in 2-3 man-weeks. Delivery of a report writer programmed by conventional techniques had been estimated to require six months and a minimum of \$30,000.

The extensibility of the META-MACHINE was demonstrated when in three days we converted an electric car performance prediction model from batch to interactive use. This model, developed by the Transportation Systems Research program at LLNL for the DOE Office of Energy Systems Research, prompts the user for different scenarios: vehicle type, battery selection, time period, driving cycle, etc., and compares the calculated result with the performance of vehicles equipped with internal combustion engines. Fifteen other models are now in use for different vehicles such as hybrids, using flywheels, batteries, and other energy storage methods. The user of the interactive models can enter his own technical values if he wishes to explore conditions that are not part of the prepared options. Validation of these models is greatly simplified by having the input data stored in individual database relations from which they can be extracted, manipulated, compared, and plotted. Results of these calculations can be saved for parameter studies and time-series analyses.

Multilingual interaction with users is possible by simple translation of the user command language, in the program database, into languages other than English. The corollary instruction strings to the underlying database management system, or to other programs, remain the same.

Different command languages can be activated by the extensible TIS database language by translating the instructions from the available INGRES command language into those of a different, local or remote, database management system or execution program. Where additional options are offered by a new system, additional prompts are appended to the program relation with the appropriate addresses and states in the first domains. In this case, the user interaction strings remain the same.

Electronic mail has been augmented by integrating the local and ARPANET directories into one mail system with many options. Automatic file transfer between word processors and TIS has been used since 1980.

Networking became operational with the incorporation of the NBS Network Access Machine software into the TIS system. Fully automated and transparent connections are made to several host machines over the ARPANET and by telephone dial-out. These include DOE/RECON, NASA/RECON, MACSYMA at MIT, NIC at SRI, LBL-UNIX, SERI, LBL-VAX, and the DOE Alternative Fuels Center in Oklahoma, among 22 other systems. To connect to DOE/RECON, for example, the user simply specifies the target by requesting: connect doerecon. The progress of the connection is displayed and requires between 7 and 45 seconds, depending on whether ARPANET or telephone lines are used. In some cases we activate at the remote host the needed resources immediately, in which case the target name is made identical to the resource, e.g.: connect macsyms.

At the present time, we offer over 64 major information resources that describe in a hierarchical manner material properties for energy storage materials, technology characterization data, systems data that are aggregates of components used in conjunction with energy storage applications, and a number of interactive models. Econometric files were added to permit market penetration studies. The results can be viewed in over 300 predefined tabular and graphical display formats.

In summary, the META-MACHINE and its implementation for the Technology Information System offers administrators and project staff a central focal point for communication and information management, more fully explained in its capabilities later on in this report. Conceptually, it functions as the basis for a stand-alone, intelligent gateway computer with capabilities of linking any user to local or remote sources of information, permitting him to extract and analyze the results, and to share them with others. To accommodate the larger traffic in communications and personal information management, we are planning to replace INGRES by installation of the directly compatible Intelligent Database Machine, IDM-500 by Britton Lee, Inc.⁵⁸ This is also expected to off-load the CPU of the PDP-11/70 and to improve throughput by at least one order of magnitude.

"Extraction of Intelligence"

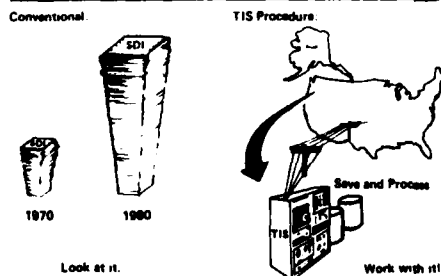
Facts, by themselves, do not convey understanding or knowledge. Their comparison with past experience, correlation with other data and/or predictions, and display in multidimensional color images, however, can be meaningful. As we are approaching the 1980s and are faced with a very large volume of factual data, it becomes increasingly necessary that we filter the data and extract new insight.

In the full-text retrieval service, NTIS has initiated a new type of automated indexing of citations with their Total Microfiche SDI Service, also known as "Selected Research in Microfiche" (SRIM). Since the fiche are being updated quarterly, the subscriber has always a reasonably up-to-date and comprehensive listing of full-length publications in his field of interest, indexed by author, subject, corporate author, contract grant number, accession report, and title--a really remarkable service making good use of the photographic storage medium. This should serve many users well in the 1980s.

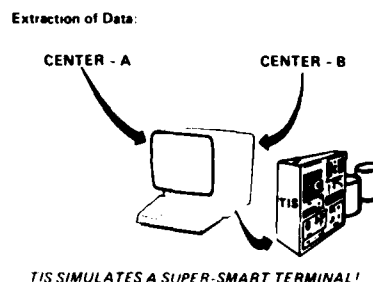
In the bibliographic online information services, both Federal and commercial citations are still being delivered to the end-user "en masse" as raw material. It was pointed out already in 1971 that much could be done with the retrieved results while they were still on the host computer. These, and related aspects, were discussed in some detail at the forum on Interactive Bibliographic Systems held at Gaithersburg, MD, October 4-5, 1971, under the chairmanship of Charles Meadow (cf. pp. 173-174):

- o Indexes by author, subject, category, etc.
- o Statistics by country, organization, etc.
- o Cross-correlations of data elements
- o Graphical display of results
- o Text analysis

POSTPROCESSING OF BIBLIOGRAPHIC INFORMATION

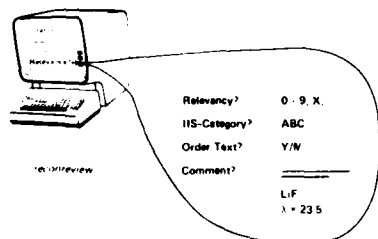


POST-PROCESSING OF NUMERIC INFORMATION:



Of these, the first is always used in books. Indexes are sorely needed in bibliographic online information retrieval. At LLNL, we use the capabilities of the Integrated Technology Information System, and offer online, interactive commands by which the above options can be carried out by the user immediately after completion of a search. The technique has been adapted especially to DOE/RECON and will be extended to other databases and information systems as required and where permissible. The user can thus create, review online, and annotate his personal or programmatic bibliography and/or library system.

ON-LINE DETERMINATION OF RELEVANT SET



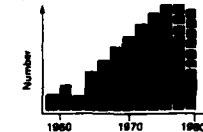
RECON ON-LINE BIBLIOGRAPHIC STATISTICS

examples

by Country

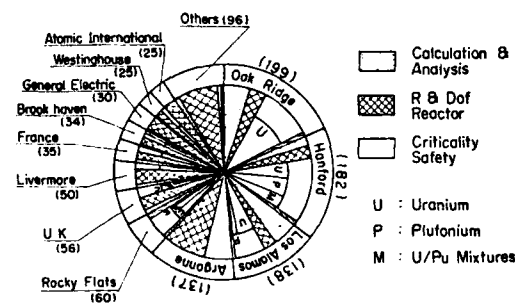
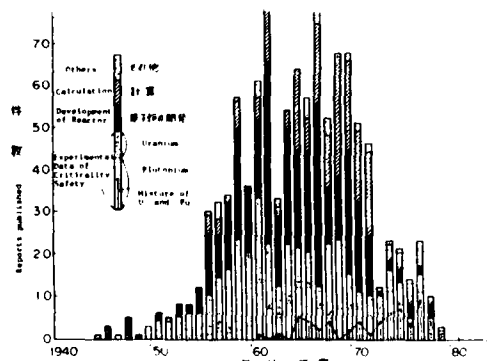
United States	n1	56%
West Germany	n2	15
France	n3	13
Russia	n4	12
Japan	n5	5
Total	N	100%

by Publication

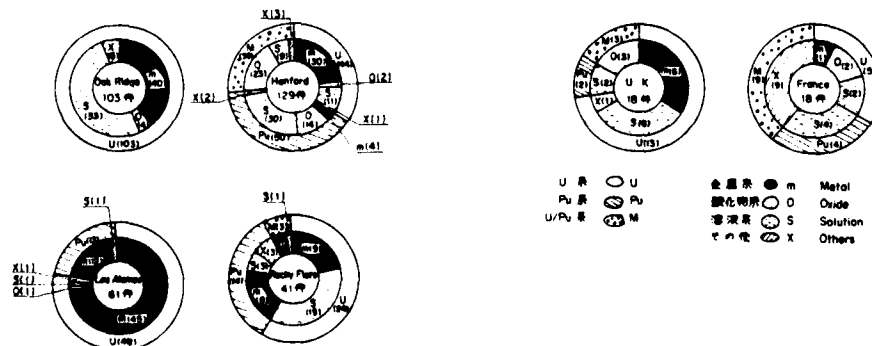


Text analysis, although not new, has some fascination because it permits us to recognize significant migration or cross-fertilization of new ideas in R&D work. Let us take for example the laser field. When one creates an authority list of all single and multiterm expressions derived from titles and abstracts, one arrives at a reasonably stable, slowly changing body of descriptive terms. Newly appearing terms can easily be set aside and used to mark citations for closer inspection. It is thus possible to find the citation where laser beams were first used to weld the retina in an eye. In other words, one can filter out those citations that somehow do not follow the common pattern of previous descriptive indexing or word usage. They contain either typographical errors or, potentially, literary pearls. Other examples quickly come to mind. The techniques for doing this type of analysis have been known since the time when inverted tables, or secondary indexes were introduced for machine-aided look-up of facts. They could be used in the 1980s as a new means of SDI service, signaling unusual and different publications.

I would like to show an additional example of higher intelligence recently extracted by Japanese technologists from a bibliography on nuclear criticality experiments published at LLNL last year.⁵⁹ This publication was released in three volumes with accurate computer-generated concordances. It served as the basis for graphical extraction of higher intelligence, e.g.:



One of the complex pie charts shows the number of reports published from each organization. Another depicts the type of experiments carried out at the different facilities, and finally, cumulative totals of experiments with different fuel enrichments are shown as a function of time. Although these illustrations were probably carried out by manual inspection of the factual data contained in the bibliography, computer programs could be written to do the same routinely.



The extraction of higher intelligence from numeric/structured data files is common practice. Numbers lend themselves directly to statistical analysis and graphical display. General and specialized programs are offered commercially or are under development at universities and Federal organizations. But, what is needed is an adaptation of some of these powerful tools for general use in a self-guided manner. As mentioned earlier, in the field of standard retrieval of bibliographic information, we see a shift from the information specialist as an intermediary to the end-consumer. The analysis and graphical display programs, however, are more complex and will require in the 1980s information specialists well familiar with the manipulation of numerical data and mathematics. The direct creation of electronic visuals (35mm slides, viewgraphs, films, etc.) by computer is the newest exploding technology for the 1980s.

An excellent example of what can be done for technologists with advanced tools is given by the Integrated Programs for Aerospace-Vehicle Design (IPAD). Early stages of the software are in use at Boeing and several other aerospace firms. The system manages the total flow of information and data for engineering and manufacturing. The National Aeronautics and Space Administration (NASA) is making major strides by sponsoring this development.

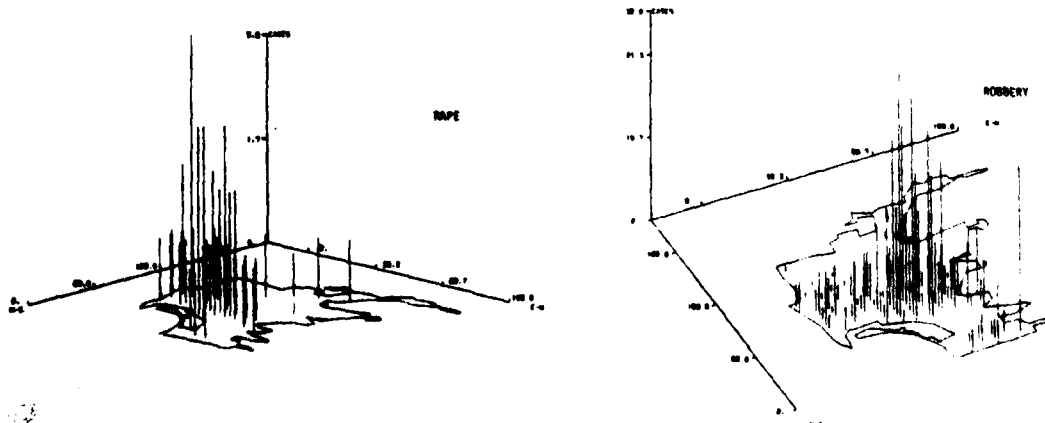
Another example is found in color graphics programs. Beautiful illustrations of double-helical DNA molecules and those of the Tomato Bushy Stunt Virus have graced the front covers of professional journals.⁶⁰ They indicate the pictorial power in store for us in years to come. At the present time, these programs are still specialized and not suitable for direct linking with databases. Illustrations are usually produced manually by one-at-a-time adaptations to recently measured parameters, or extractions from an existing database. What is needed here, too, is a user-oriented tool by which anyone with access to the CIS databases, for example, could search, retrieve, display, and zoom in on the molecular structures of interest to him. A number of these sophisticated tools are in the public domain and will hopefully become available in simplified form as well.



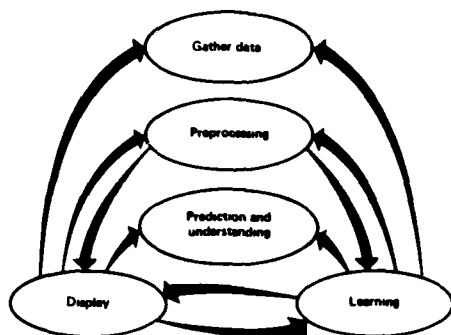
Cutaway view showing inner surface of protein coat of Tomato Bushy Stunt Virus, as represented by computer graphics system under development at National Resource for Computation in Chemistry (NRCC) and LLNL. Each large sphere represents one protein molecule, small spheres are "linking arms" of the protein molecules. Close-up view focuses on three of the protein molecules in the TBS virus, showing arms that link molecules.

I would like to conclude this section with a remark about pattern recognition. This powerful fact retrieval technique has been used primarily with numeric data, but there is no reason why it could not be used with structured, coded data or even with text. The following examples may serve this purpose:

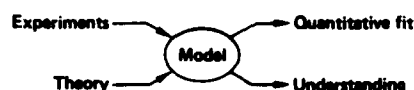
The San Diego Police Department received a grant in 1975 from the Law Enforcement Assistant Administration (ARJIS) to plan for a computerized data processing system that could serve all of San Diego County. One of the goals was the implementation of an automated crime analysis capability. The Lawrence Livermore National Laboratory was requested to assist ARJIS project personnel under auspices of the National Technology Transfer Program. The results were impressive. Pattern recognition with the LLNL-developed PATTERN program could identify correctly those crimes where the success of investigation and conviction was highest. Such an approach can be used for similar multivariate problems that are difficult to solve by statistical or conventional means. In another case, the same pattern recognition program was applied successfully to the primary data describing elements on a wall chart. It predicted correctly the acidic, basic, and graduated amphoteric characteristics of the elements. Pattern recognition has particular importance for work of unusual complexity and offers powerful solutions to problems with limited manpower and budgets.



AT LLL WE EMPHASIZE THE ITERATIVE AND INTERACTIVE NATURE OF PATTERN RECOGNITION ANALYSIS -



MODELING



- Theory is often insufficient
- Experimental measurements
- Insufficient data for statistical modeling

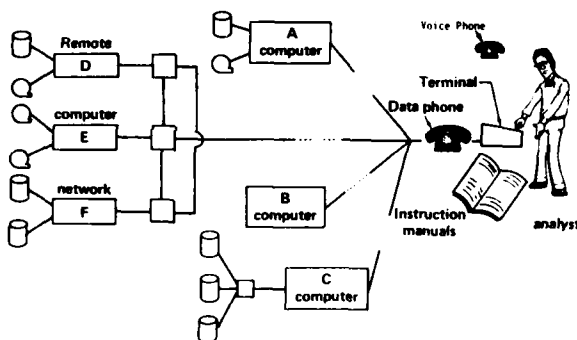
Pattern recognition techniques can be used to construct semi-empirical models

References: C.F. Bender et al.,
Pattern Recognition, Jour. Am. Chem. Soc. 94, 5632, 1972.
Pattern Recognition & Crime Analysis, UCID-17224, 1976.

"The Intelligent Gateway Computer"

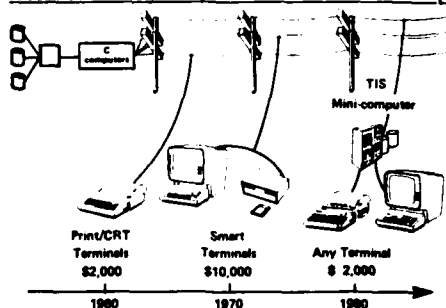
Local office networks are starting to be interconnected with computers, remote information centers, and other resources. The results are not always simple or elegant, but they work and are prompted by the availability of networks and satellites which are removing distance-dependent communication costs. These developments magnify the opportunities for expanding available information resources and computer power enormously. However, to

MANUAL MODE OF OPERATION

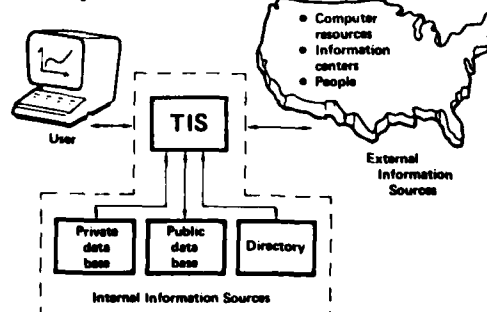


circumvent dissimilarities among different resources and communication protocols, a coordinated approach is needed. Ideally, it should permit a group of users to interact with each other, and with the rest of the world, in a cost-effective and practical manner.

EVOLUTION OF COMPUTER ACCESS TERMINALS



Basic Configuration



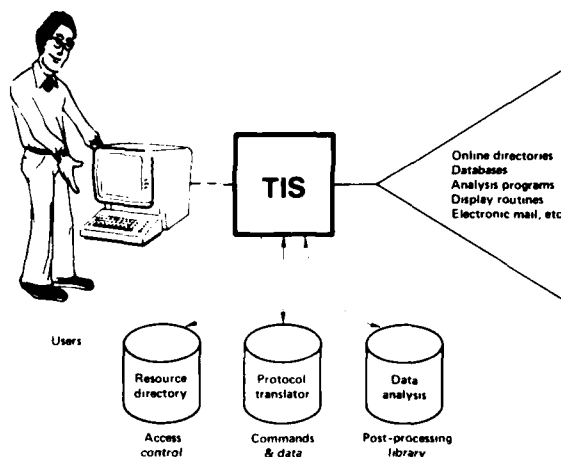
The trend toward this direction has been apparent for some time. Computer terminals developed gradually from mere keyboards to being imbued with microprocessor brains and memories. The communication capabilities of these intelligent terminals, however, are usually intended to connect only to one, or to a few, remote computers or resources. Terminals of this type are offered for \$15K to \$25K, where some would be considered to be in the class

of minicomputers. In most cases, the manufacturers provide capabilities that are upward compatible only within their line of hardware. This prompts each owner and user of such an intelligent terminal to "upgrade" with even greater investments in time and money without necessarily coming closer to a more flexible and automated solution.

Common communication gateway computers also provide only a partial solution. As a rule, they are not very intelligent. They connect and may exert some control and perhaps do some accounting, but they do not provide translation of incompatible protocols among electronic text processors; they do not permit users to determine their capabilities or allow extraction and saving of information and data from other sources.

What is needed is a stand-alone, intelligent gateway computer. It should contain a master index to the available resources, connect authorized users automatically, translate protocols and formats, permit the aggregation of reasonable amounts of extracted information and data, and offer a resident library of software tools by which these data can be post-processed, analyzed, and displayed.

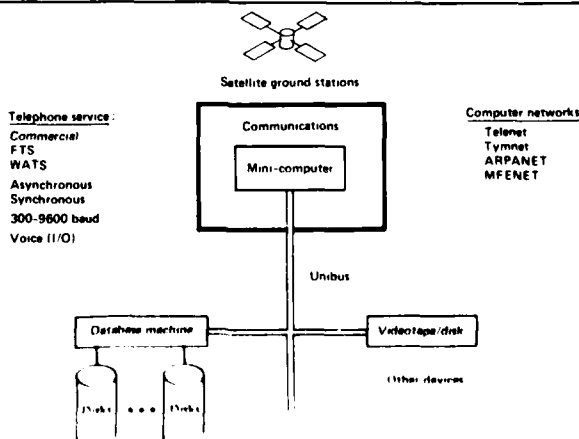
An Intelligent Gateway Computer



At LLNL, we were confronted with a similar problem in 1975 when the U.S. Department of Transportation asked us to study the concept and implementation of a "Transaction Controller" that would permit their analysts to interact with some 26 different computer centers where the statistics of passenger and cargo traffic by ship, air, trains, and trucks are kept. Our work evolved into the design and implementation of the META-MACHINE user interface for the Technology Information System, described previously. Presently, we are generalizing our experience in this field with the concept of a stand-alone, portable system.

The "Intelligent Gateway Computers" would serve a group of users who could retain their old terminals since the intelligence can now be made to reside in a time-shared, interactive minicomputer. A possible configuration might be a PDP-VAX-750 machine, coupled with an IDM-500 back-end database machine, and a versatile communications front for automated dial-out, dial-in, and network access. Our development work at LLNL points toward this direction and uses the META-MACHINE as the extensible and flexible interface:

CONCEPTUAL VIEW OF A FUTURE "TIS" INTELLIGENT GATEWAY COMPUTER



4. FACT RETRIEVAL IN THREE DIMENSIONS AS A FUNCTION OF TIME

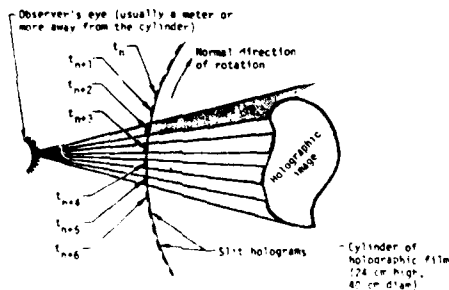
This is probably the ultimate goal of information management. But, as with speech and books, the technology advances first to a state where it can faithfully store and duplicate the information. Management of information, and extraction of higher levels of intelligence, follows later on. Here, I would like to draw your attention to two unusual three-dimensional display methods, those of holography and stereo-vision. Both were recently reviewed by Dr. Donald L. Vickers who is leading the graphics group at the Lawrence Livermore National Laboratory. The technical details and illustrations are extracts of his presentation⁶¹ and should give you an immediate insight to the potential of this upcoming technology for your future work.

"Crystallographers, for example, are encumbered with more than their share of the age-old-problem of trying to perceive three-dimensional information. Consequently, they are more painfully aware than most of the irony of living in a three-dimensional world while having to communicate with fewer than three dimensions. Even the computer, with its great speed and ability to solve problems based on data with many dimensions, usually ignores the potential of communicating with the user in at least three dimensions when it comes to graphically presenting the results. Three-dimensional, computer-stored information nearly always is transformed into two dimensions and then plotted on paper, film, or the face of a cathode ray tube (CRT). There are, however, several displays on which three-dimensional computer data actually appear three-dimensional and stereoscopic. Two such displays are the integral hologram and the head-mounted display.

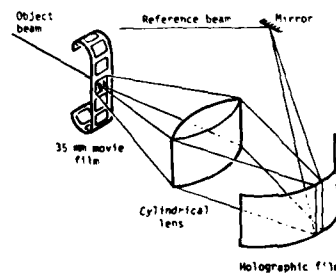
"Integral Holography"

Integral holography is several steps beyond the holography most are familiar with. Many scientists have seen a standard transmission hologram and the three-dimensional image which it forms when illuminated with the monochromatic light of a laser. Many have also seen the cylinder hologram, a cylinder of processed holographic film which, when illuminated with a laser, recreates a three-dimensional image inside the cylinder. As the cylinder is rotated, the object inside also appears to rotate and, thus, one may view both the back and front of the holographic image. Now, try to imagine a cylinder hologram that produces a holographic image which not only rotates as the cylinder is rotated, but which also moves or deforms -- this is an integral hologram. (Refer to figures on the next page.

But even more amazing than a time-varying holographic image is that the source of illumination for viewing is an ordinary incandescent light bulb and not a laser. An integral hologram is not made from a single exposure as it is in an ordinary cylinder hologram, rather, it is made up of 2160 individual slit holograms which are integrated by the observer's eye to form what appears to be a single holographic image. The 2160 slit holograms are made from consecutive frames of a 1080-frame 35-mm black and white movie film by a process described in the following paragraph. The holographic film containing the slit holograms is taped to the outside of a plastic cylinder about 40 cm in diameter such that each slit hologram subtends $1/60^\circ$ of arc on the cylinder. As the cylinder is rotated, successive slit holograms come into view, giving to the holographic image an illusion of motion much like that in the 35-mm movie from which the integral hologram was made. The 1080 frames correspond to a 45-second movie shown at a rate of 24 frames per second.



Schematic view from above showing how various slit holograms, though separate in time, form a single "integrated" holographic image. When the cylinder rotates, the eye does not notice any separation between the slit holograms.



Schematic diagram of the main elements used by the Multiplex Company for making integral holograms.

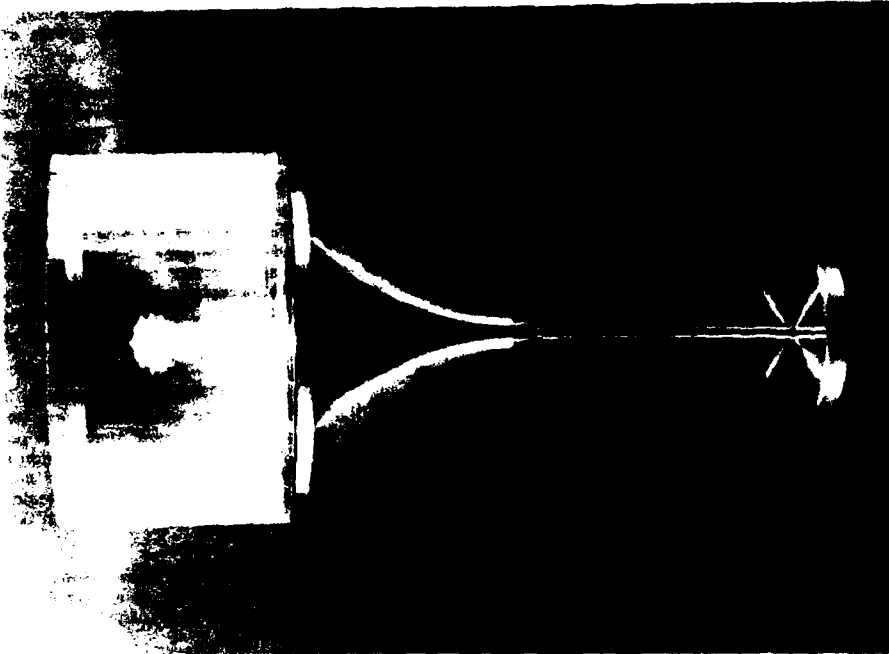
The slit holograms are 24 cm high and 0.7 cm wide on the 1.25-m-long holographic film. They are made by shining a 7-mW HeNe laser through one frame of the 35-mm movie at a time, passing the resultant beam through a large cylindrical lens, and combining it with a reference beam. The interference patterns thereby generated are captured on holographic film which is advanced about 0.5 mm between exposures, thus causing some overlap among adjacent slit holograms. In the "standard" grade integral hologram, the apparent resolution of the holographic image is improved by exposing each frame of the 35-mm movie twice. This is why the 1080 frames of the 35-mm movie result in 2160 slit holograms. A coarser-looking "proof" grade hologram can be made by advancing the holographic film 1 mm and exposing each 35-mm frame once.

Two factors combine to allow the use of white light rather than laser light for illumination of the hologram. First, a cylindrical lens instead of a diffusion screen is used in the exposing process. Second, the subject of the hologram is a two-dimensional piece of 35-mm film and does not require a laser to "unlock" the third dimension. In fact, if monochromatic light were used one would not see the whole image but just a small hoop-shaped band of it and the band would move up and down showing different parts of the holographic image as one raised and lowered one's head. As it is, the holographic film acts as a diffraction grating for the illuminating white light so that the holographic image one sees contains the colors of the rainbow, ranging from red at the top to violet at the bottom. As one moves up and down while looking at the holographic image the rainbow of colors appears to shift. Two restrictions are imposed on the nature of the illuminating white light: the bulb must be unfrosted, and the filament must be as near as practical to a single vertical line.

To date, the Lawrence Livermore National Laboratory (LLNL) has produced five computer-generated movies of scientific interest from which integral holograms have been made.⁶² One shows a nonrotating disk which, as the



Viewing stand showing the Time-Resolved X-ray Integral hologram. This two-meter-high stand is on continual display in the lobby of one of the main buildings at Lawrence Livermore National Laboratory. The X-ray hologram was shown at the NATO/AGARD meeting.



Integral hologram showing the calcite crystal structure displayed in a portable viewing stand. The stand, about 40 cm in diameter, uses an ordinary incandescent light bulb for illumination. This stand and hologram were displayed at the NATO/AGARD meeting.

cylinder is rotated, warps according to the phase change in the beam of one of our large lasers. Another shows a rotating sheet which "grows into a mountain," the shape and altitude of which indicate the density of X rays emitted from a laser-bombarded target. Two of the five holograms are of direct interest to crystallographers. The first shows a rotating tetraglycine molecule which, as it rotates, changes from a "ball-and-stick" to a "filled-space" model. The second represents a rotating calcite crystal showing first the unit cell, then adding a cleavage rhombohedron, and finally showing the free rotation of the oxygen atoms about the carbon atoms within the carbonate group, as it actually occurs at extremely high temperatures. Refer to figures on preceding page.

The concepts of integral holography are not new. Both the ideas of making composite holograms and of making, from a collection of two-dimensional perspectives (movies), holograms which can be viewed with white light have previously been published. Also in the literature are articles about how to make full-color holograms. What is new is the working combination of these concepts and the commercial availability of the integral hologram such that any interested researcher may take advantage of it.

Applications of integral holography to the areas of advertising and publicity are obvious. The integral hologram is also potentially very useful to both scientific education and research. Imagine the benefit of showing students not just a rotating three-dimensional model of a molecule, but a rotating three-dimensional holographic model that also shows how molecules and atoms combine during a chemical reaction. Already integral holograms have been used with resounding success as visual aids for technical papers in the area of chemistry, crystallography, and computer science. When used as an aid to help the researcher better understand the structure of some molecule or crystal he is working with, the integral hologram is more accurate, more versatile, less expensive, and requires much less effort to create than most complicated ball-and-stick models.

Up to now the integral holograms we have dealt with have been limited to 360° views. (Actually the tetraglycine and calcite integral holograms may be considered to have 720° of view since the molecules rotate twice for every rotation of the holographic cylinder.) Researchers, using a process which requires a monochromatic point source for illumination,⁶³ made an experimental integral hologram on a piece of holographic film 70 mm wide and about 7 m long which they spiraled around a plastic cylinder and were thus able to view a much longer movie sequence. When the process of making an integral hologram from a movie becomes more automated, it should be possible to use source and take-up reels which move a full-length movie's worth of holographic film around a rotating plastic viewing cylinder.

Fact retrieval by means of integral holography for most of us may still be some time in the future. However, we recognize that these advanced capabilities are coming up and have to be linked with the data in a user-oriented environment. This is where the producers of data and information specialists come in. Currently, there is probably only one place in the world with the facility for making a white-light integral hologram.

"Stereo Vision of Facts"

We just learned how computer-generated integral holograms permit us to view complex data as three-dimensional movies. But we remained passive observers of a virtual, tantalizing imagery. To remember these images, to explore the meaning of their valleys and peaks, we still would have to use our minds to ponder their implications from different points of view. Would it not be nice and informative to be able to walk into the data, being able to touch them, to "feel" their irregularities, and to smoothen them by hand. We may have done some of it in an abstract sense mathematically, but must have envied the sculptor who can project the images of his inward eye into reality and mold it by hand.

The technology to do this in information management is available. It has been explored during the seventies. I am bringing it here to your attention because technical difficulties and costs that may have prevented its widespread application in earlier years have decreased with the advent of inexpensive, powerful microcomputers that feed on data from larger stores. Stereo-vision of facts is made possible by a head-mounted, three-dimensional viewer. It was first built by Dr. Ivan E. Sutherland at Harvard College.⁶⁴

The display itself uses refreshed CRT technology, but, unlike most CRT displays, this one is worn on the head like a pair of spectacles. Pictures drawn by a computer on the two CRTs are presented to the person wearing the spectacles, the observer, as a virtual image which appears to be made of glowing wire and is superimposed on the observer's field of view in such a way that it seems to float in space. The viewer and a virtual map of the United States are shown on the next page.

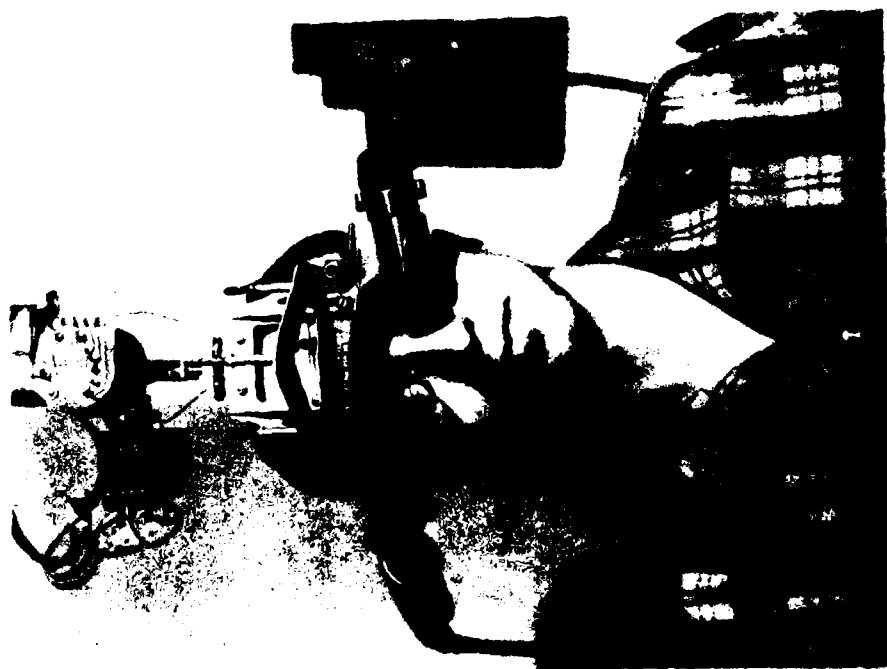
The computer-generated, or "synthetic," objects can be programmed to remain stationary as the observer walks around, between, or even through them. They can also be made to change in time as the observer stands still or as he evokes a response by his body motion or controls. Both a wand and light-studded gloves have been used to reach out and interact with the synthetic objects, allowing an observer to touch, connect, deform, erase, and even to create them by "drawing" in space. The unique three-dimensional environment created by the head-mounted display comes from the smooth teamwork of many special pieces of hardware. Refer to Figures.

The head set which actually presents the synthetic object(s) to an observer has two 2-cm-diameter CRTs mounted where the temple pieces would be located in normal eyeglasses. A picture drawn on the left CRT, for instance, is reflected off a mirror, through a lens, onto a half-silvered prism, and from there into the observer's left eye. The prism allows the observed to see the real objects in the surrounding room plus the virtual images or synthetic objects drawn by the computer. Stereoscopic viewing is possible by sending different pictures to the right and left eyes. In order for the computer to make a synthetic object appear stationary as the observer moves about, the computer must monitor the position and orientation of the observer's head. This is done with a counterbalanced head-position sensor which consists of a 2-m-long telescoping tube attached through universal joints to both the head set and pivotal reference point on the ceiling. The cyclic chain of events from the reading of the head-position sensors to the drawing of the vectors on the CRTs, takes place at the rate of at least 20 times a second in order to avoid flicker or jerking on the CRT. Future improvements may well use nonmechanical sensors to establish the head position of the viewer. The interaction of different components is shown below.

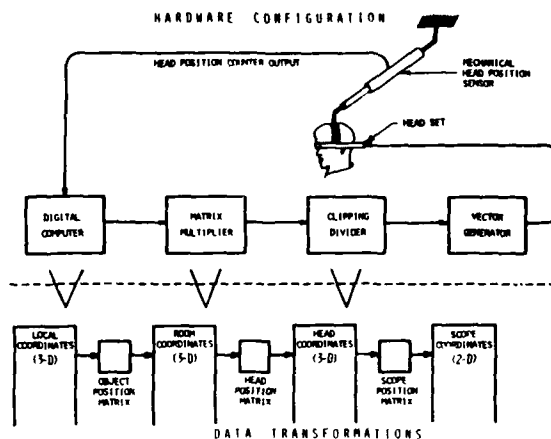
* The Multiplex Company of San Francisco, which has applied for a patent on their process. You may wish to address your inquiries to them or to Don Vickers, LLNL Graphics Group, L-73, Lawrence Livermore National Laboratory, P.O. Box 808, Livermore, CA 94550, USA.



USA map photographed from different angles through a head set eyepiece. The map consists of 649 line segments and was programmed to appear about 1 m wide and 0.6 m high. Data for the map was supplied by the Evans & Sutherland Computer Corporation. Only one view is shown here.



Observer wearing the head set and harness of the head-mounted display. The lower part of the head position sensor shaft and the encoders for the lower universal joint are visible at the top of the figure.



The wand, the most frequently used device for interaction with the three-dimensional synthetic objects, is equipped with four buttons, a switch, and a potentiometer. These signaling devices allow the observer to "tell" the computer whether he is drawing, erasing, or extracting data. The computer tracks the position of the wand by monitoring the length of line attached to take-up reels on the ceiling. These three spring-loaded lines also serve to counterbalance the weight of the wand and the cord attached to it. (Again, wireless position sensors can readily be envisioned to provide mobility to the user.)

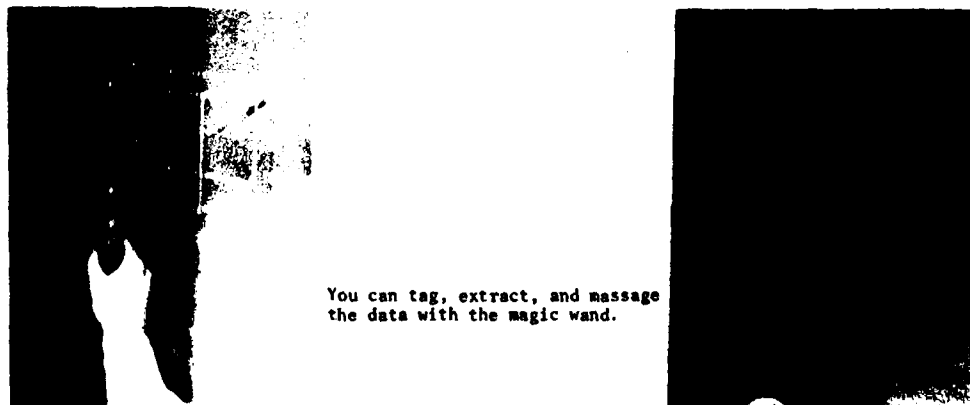
It was quickly evident that four buttons were insufficient to do all the things one might like to do with such a wand, so a wall chart divided into quadrants was designed to multiply the effect of the buttons. Each quadrant of the chart corresponds to a different mode of operation, and each mode redefines the meaning of the four buttons. Of course, only an observer wearing the head set could see the confirmation signaled by the computer to the viewer in the form of a virtual cross. Thus the wand allowed interaction not only with the synthetic objects, but also with any other objects in the room which were "known" to the computer.

Though the head-mounted display was built at Harvard College, it was taken to the University of Utah by Professor Sutherland just a week or two after its completion. At the University of Utah it served as a research tool for several graduate students. Currently, however, it is inoperable and, though interest persists, funds are lacking. Clearly, the head-mounted display is a prototype. It is not as yet a resource available to present-day crystallographers or to scientists of other disciplines.

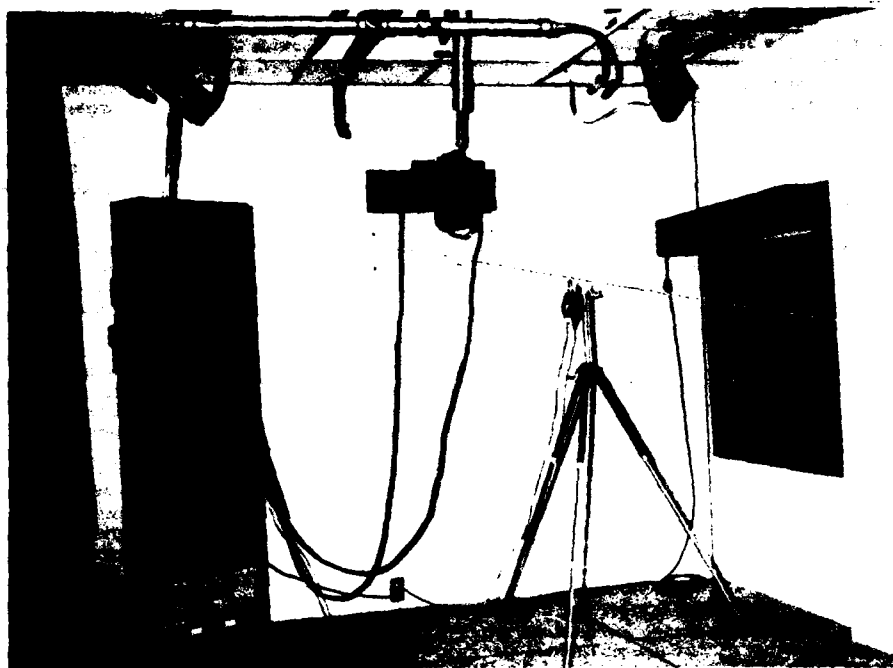
Nonetheless, the head-mounted display has been used with modest success for looking at three-dimensional mathematical functions such as a 3-4-5 Lissajous Figure, for analyzing three-dimensional electrocardiogram data, and for simulating arterial structure (one was able to walk through simulated arteries). It has been used to "trace" three-dimensional objects, to design three-dimensional, mathematically-defined surfaces and to simulate what a blind person would "see" if an electrode array were implanted in his visual cortex. To our knowledge, it never has been used to look at molecular or crystallographic data.

It is not difficult to imagine a next-generation head-mounted display system which would show colored synthetic objects as solids rather than as outlined drawings. Such a display might be equipped with remote head-position sensing and with telemetered communication both from the wand and to the CRTs, leaving the observer completely free to move around with no wires or harnesses to worry about. Such a display might be connected to a computer which is processing x-ray diffraction data and could be used to look online at a crystal structure model as it emerges from that data. The wand could be used to single out certain atoms in the synthetic image, or to make the image shrink or grow, allowing the observer to walk around it or even inside it.

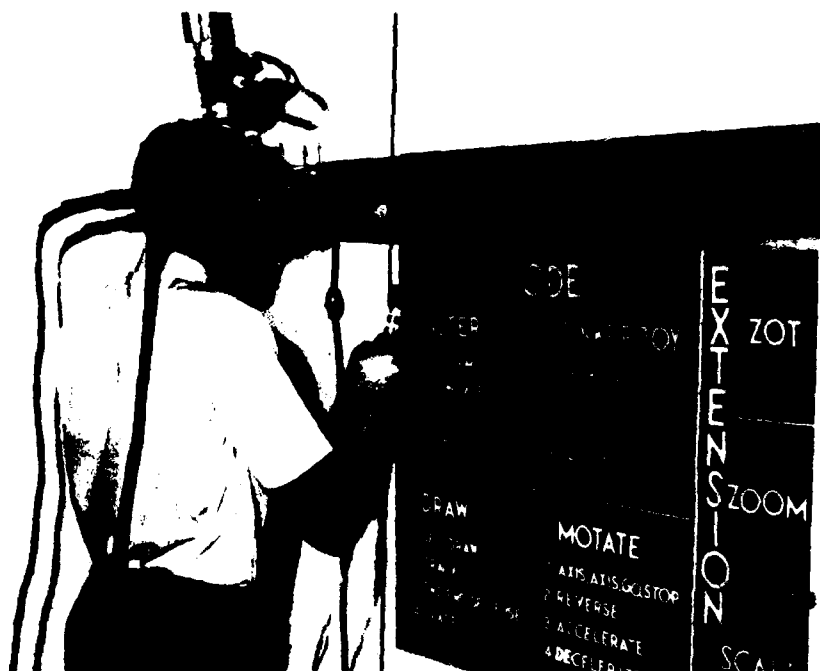
The qualities that make the head-mounted display different from any other three-dimensional graphical I/O device are its abilities: (1) to produce three-dimensional images which are constantly updated under computer control and which change their view as the observer walks naturally about, just as real objects do, and (2) to superimpose a three-dimensional synthetic object on real three-dimensional objects. The data control wand adds the power: (1) to interact with these three-dimensional objects using natural pointing motions rather than by turning knobs or pushing buttons, and (2) to interact with real objects in the surrounding room. And, in spite of being limited to "wire-frame" drawings, the synthetic objects look startlingly real! When this comes about, the finding of facts may be as simple as picking up a synthetic book from a shelf, checking its computer-based index, and reading its virtual pages with all the quickness of the human mind, and projecting their data into space for a personal look and inspection. Computer-aided fact retrieval will then have become an extension of our natural environment."



You can tag, extract, and massage the data with the magic wand.



View of computer head-mounted, three-dimensional viewer, hand-held wand for data manipulation, and control board on wall.



Observer pointing with the wand at the wall chart. The head mounted display and wand allowed user interaction not only with the computer-drawn objects floating in space, but also with real objects such as the chart.

5. THE INTEGRATED "TECHNOLOGY INFORMATION SYSTEM" AT LLNL

Capabilities of the Technology Information System (TIS) provide nationwide bibliographic and numeric database management, interactive modeling, electronic communications, and distributed networking. These capabilities are self-guided and are used successfully by those not intimately familiar with computers. The description of TIS is given here as an example of an operational, intelligent gateway computer, expected to serve technologists throughout the 1980s.^{65,66}

TIS is a new-generation, dedicated information machine. Programmatic information is kept on TIS. When additional information or numeric data are needed, TIS connects to other information centers, in an automated and controlled manner. Authorized users simply specify the target name of the desired resource.

In addition, since much of the daily work in R&D is being documented on electronic word processors, we established the capability of linking with several of these machines for transfer of information and data to and from TIS. Translation of formats to WANG, LEXITRON, and QYX word processors is carried out by TIS as required. Now that commercial hardware/software have started to appear on the market, we are planning their procurement to free our resources for other areas.

Analysis, synthesis, and post-processing of information and data are needed to speed up progress, increase productivity, and transfer technology. TIS gives this capability to each user. You, the user, can define and create your own data files, reports, graphics, and communications by activating self-guiding routines. Initially, the results of your work belong to you alone. It requires a permit command to share the data or displays with someone else, a group of co-workers, or to release them for general use.

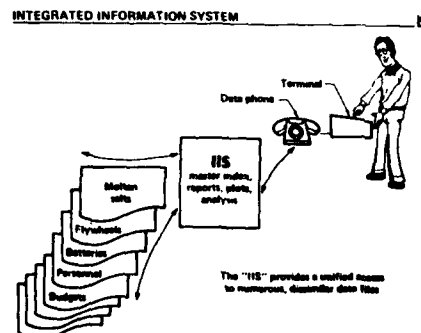
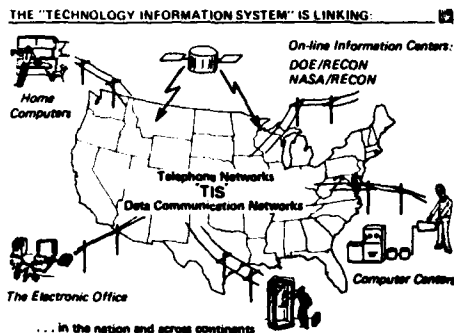
The system is accessible from any telephone at 300 or 1200 bps, over the ARPA computer network, and soon also over the worldwide TELENET/TYMNET system. FTS and WATS lines are provided for cost-effective use of communications and convenience. Here, I wish to highlight the major capabilities in database management, modeling, and communications. I hope that some of our experience may be useful to you as you are planning your integrated information systems.

The Technology Information System has been supported by the DOE Office for Energy Systems Research (DOE/ESR). The TIS user community now includes, principally, the staff of the Transportation Systems Research program at LLNL, and that of the Seasonal Thermal Energy Storage program at Battelle, PNL, as well as the DOE/ESR staff. There are about 170 authorized users throughout the country. Electronic communications and the automated access to other information centers is available to all users. This capability is used extensively by the Interagency Information Exchange committee and the DOE Technical Information Center. We are beginning to prototype an Integrated Information Network.

"Database Management Capabilities"

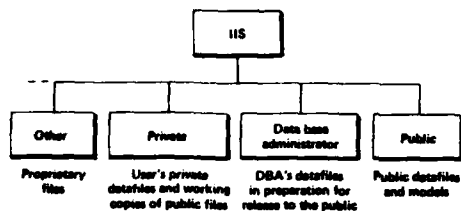
Information is the total of textual and numeric data displayed in a meaningful manner. Most systems specialize in one or the other. Also, most database management systems require a computer programmer or analyst to define the schema of a new database and to load it. Then, when this is done, the database is turned over to the user for retrieval and updating. When special features are required, e.g., more complex reports or graphical output, the services of a programmer are again needed.

The Technology Information System (TIS) offers the traditional database management procedures, but, in addition, TIS has the capability of direct database management by its users without programmer intervention. This permits the use of TIS as an extension of the yellow note pad, or desk calculator. Thus we distinguish on TIS two categories of databases, public and private.



The information in these databases is displayed in a hierarchical manner and can be selected with simple specification of an "Option Number," not unlike what you find in some of the popular word processors. However, a database on TIS is a collection of programmatic resources: data files, models, and electronic communications are all different options of the same database and can be tailored to individual programmatic requirements. Six major databases are on the system now. We name especially the database established by the Transportation Systems Research (TSR) program⁶⁷ for DOE/ESR, the STES database, and that which contains the installation and technical specifications of thousands of expensive pieces of optics for the SHIVA/NOVA laser fusion program at LLNL.⁶⁸ An extract from the display of the TSR database follows:

DOE/ACT TECHNOLOGY INFORMATION SYSTEM



0 AVAILABLE RESOURCES

1	NEWS
2	MATERIAL PROPERTIES
3	ENERGY STORAGE COMPONENTS
4	ENERGY STORAGE SYSTEMS
5	INTERACTIVE VEHICLE MODELS
6	ECONOMIC ANALYSES OF ENERGY STORAGE TECHNOLOGY
7	TRANSPORTATION STATISTICS
8	TRANSPORTATION SYSTEMS RESEARCH
9	NATIONAL AND WORLDWIDE ENERGY DATA
10	NATIONAL ECONOMIC DATA
11	BIBLIOGRAPHIES
12	DOE/STOR ADMINISTRATIVE INFORMATION
13	INTEGRATED COMPUTER RESOURCES
14	UTILITY COMMANDS
15	MAIL
16	HELP

The public databases are intended for general use in support of a particular program. Their information content can be viewed, used, and extracted as required. Temporary changes to these data can be done for display or for ad hoc exploratory calculations by any user. When such changes are made, they are annunciated in the input record documenting a report or model run. Permanent changes can only be initiated by the originators of the data with the help of the Database Administrator.

In the private database, we offer the additional capabilities of database creation. The *create* command, starts a self-guided routine that permits you to establish a hierarchical index for information in your own database system. You can specify and name the data files, and are prompted to describe each data field, indicating whether it will be used for textual data, integer data, or floating-point data; you are then asked to name the units of measurement and select an acronym by which you may wish to refer to the data field in the future as an equally valid name for the corresponding Option Number.

When these self-guided definitions are completed, data can be entered key-to-disk, from menu-driven forms that flash on the cathode ray screen. These display formats can be activated by initiation of the self-guided *makeform* routines and can be called into action as needed by name. In several cases we have had good results with data input of this type cross-country by secretarial help. The data fields are explicitly called out on the screen. When text characters are inadvertently entered into a numeric data field, the terminal keyboard locks and signals the error for immediate correction before proceeding. The *update* command is used to append, replace and replicate data. It provides help instructions for the searching of erroneous records which can then be corrected in a systematic manner. Magnetic tapes are used for input when larger volumes of data are involved. Data can also be transferred over telephone lines or over the ARPA computer network. In the latter case, we can accept data at an effective transmission rate of 36,000 bps.

"TIS" OPERATIONAL RESOURCES:

Data bases:	64	Technical and economic datafiles
Models:	14	Electric-vehicle models
	1	STES Models
Communications:	64	Communication ports
		300-baud telephone lines (Comm & FTS)
		1200-baud telephone lines (Comm & FTS)
		300/1200-baud WATS lines
		ARPANET
		Telephone dial-out
		Dedicated lines, on site at LLL
		VOTRAX Voice synthesizer
		TYMNET/TELNET
Other computers:	22	Automated links to other computers/centers:
		DOE/RECON, NASA/RECON, DOE alternative
		fuels (OK), SERI, DOT/TSC, LBL-SEEDIS,
		MIT-MACSYMA, Worldnews (NYT, UP), etc.

"TIS" GIVES YOU THE CAPABILITY TO CREATE YOUR OWN INFORMATION SYSTEM:

Give the command:

* Create

Index?

Datafile?

Title.....

Data fields?

description

type: c, i, f

length?

abbreviation

units

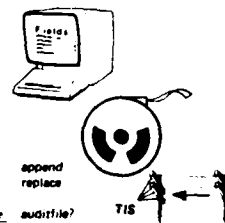
* Makeform

fields?

title?

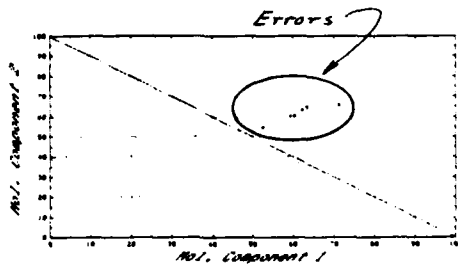
* Update

auditfile?



The display and extraction of information or numeric data can be carried out in two ways. First, each public data file comes equipped with one or several preferred display formats, also referred to as reports. These can be activated by name and provide the option to specify Boolean logic for the selection of those records that satisfy numeric and/or textual criteria. Reports can be graphs or tables. You can choose those that fit your terminal. Second, you can create your own reports by initiation of the *print* or *plot* commands. These routines guide you to indicate the data file for which the report is intended, the datafields to be printed, summed, labeled, ordered, footnoted, etc. Plots can be seen in black and white on Hewlett-Packard 2648 terminals, or in color on HP 7221 X-Y color plotters. The latter can be used to make viewgraphs directly. These display patterns can be combined with text for reports which are activated by name as an automated *sequence* of commands. A device-independent interface is being prepared for graphics display on terminals from other manufacturers. In the example below, which shows a scatterplot of the 6200 Eutectic salts, created on TIS from the magnetic print tape to the corresponding NBS publication,⁶⁹ we discovered 7 salts where the percentages of constituent materials exceeded 100%. These errors were brought to the attention of NBS/OSRD. This clearly illustrates the power of database management for Information Analysis Centers.

When needed, numeric data can be extracted for later use. They are prepared through the *print* command and then saved in separate files which serve as direct input to models or as inclusions in electronic mail.



Type an Option Number, a Command, or "step"
Report Eutectic where mol1 = mol2 = 100

Molten Salt Eutectic Report where mol1 = mol2 = 100

MOLTEN SALT EUTECTIC DATA

U.S. SALT	MOL. %	TEMP. (C)	REFERENCES
208 NaCl2-HaCl-TaCl5	83.1-37.2-1.52	78.00	23
366 NaCl2-HaCl2	62-33	114.39	159
450 NaCl2-HaCl	71-25.5		680
2881 CaCl2-HaCl	52.5-55		42 59
			62 83
2888 CaCl2-LiCl	82-84	300.00	255 259
4385 NaCl-TaF5	83-84	100.00	114
4684 KF-HaF	59.5-60		8 51
			61 45
			12 179

Type an Option Number, a Command, or "step"

Every report, graph, and sequence of presentations can be used initially only by the creator. A positive permit command is required to share the information with someone else, a group of co-workers, or the user community as a whole. Both in the public and private databases, the user is shown the availability of only those data files, report formats, and display patterns to which he has access.

For those who like to venture into more complex work with data files and reporting, many powerful UNIX utility routines are available. All routines are documented online.

Help is available online for most programs. Commands with many parameter options give help during initiation by the user prior to execution. You may type "help" at each step to receive guidance for the next question to be answered. We offer also online tutorials. This is described later when the link command is discussed under TIS communications.

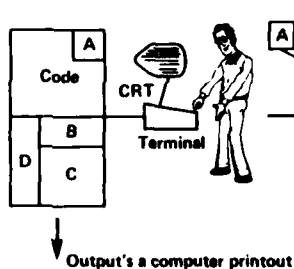
"Modeling"

The execution of simulation models for performance prediction of energy storage systems, or for technical and economic analysis, can be carried out interactively or in the batch mode, in three ways as shown below. We prefer the first mode, which permits separation of data files from the models with inherent advantages.

MODELING

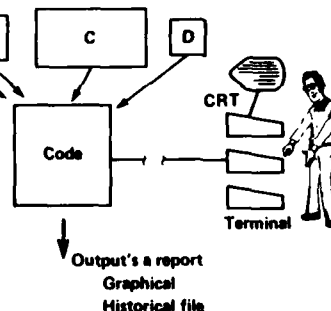
Old method:

- Batch processing
- Few users
- Data and Program intertwined



New method:

- Interactive
- Many users
- Data and Program separated



- o The model may reside on TIS, which controls its input and output.
- o The model may reside in TIS, and be activated by TIS, but prompting is carried out under model control. This is usually the case for imported models.
- o The model may reside on another computer elsewhere in the country, but is controlled by TIS with regard to input or output.

The first method is well represented by the group of models developed at LLNL for performance prediction of electric and hybrid vehicles, powered on various energy storage devices. Originally, these models were used in the batch mode and the code was intertwined with input data. Their use was difficult, limited by computer printouts, and best left to those familiar with their intricacies. To make them interactive, and to prepare them for use by others, the modelers prepared succinct statements that describe the purpose, assumptions, methodology, and limitations of each

model. These descriptions were then used for their online documentation in TIS. Next, the modelers wrote the interactive script that describes each parameter, assumptions, modeling techniques, and limitations. During execution of the model, the user is prompted to select categories and input parameters in a self-guided manner.^{70,71}

Type an Option Number, a Command, or "stop".
*5

Type an Option Number, a Command, or "stop".
*5.3

5	Interactive Vehicle Models
5.1	Electric Vehicle Data
5.2	Standard Driving Cycles
5.3	Electric Vehicle with Optimized Battery
5.4	Battery-flywheel Vehicle
5.5	EXXON Energy Storage Evaluation Model
5.6	JPLEV (Elect. Veh. Model)
5.7	Aluminum/Air Battery Vehicle Model
5.8	CARR Electric Vehicle Model

Type an Option Number, a Command, or "stop".
*

5.3	Electric Vehicle with Optimized Battery
5.3.1	Model Purpose and Assumptions
5.3.2	The Modeling Technique
5.3.3	Basic Vehicle Types
5.3.4	Basic Performance Levels
5.3.5	Basic Time Periods
5.3.6	Basic Confidence Levels
5.3.7	Battery Descriptions
5.3.8	Driving Cycles
5.3.9	Running the Model

Type an Option Number, a Command, or "stop".
*

The descriptions and the interactive script for each model were then placed in a small data file and were made part of the overall Transportation Systems Research database. They provide, in this manner, dynamic, up-to-date documentation. The potential user can familiarize himself with any aspect of the model background by selective reference to the appropriate Option Numbers, which describe the origin, purpose, techniques, limitations, and input data files. When he is ready to run the model, he is prompted to select the required parameter categories, or to give his own values. When all answers have been received by TIS, the data are extracted from the individual data files and presented for viewing and confirmation. Ad hoc changes can be made at this time. They affect the run, but not the content of the public database. Following execution in real time, the results are presented in tabular form or as graphical output. We have also devised efficient interactive input methods which the user may wish to use for repeated execution of models.

INPUT TO MODEL

THIS COMPUTER RUN WILL START WITH A BASELINE
ICE VEHICLE DESCRIBED AS FOLLOWS:

2 passenger car (vt*2) limited performance (pi*3)
The projected use is for the federal urban cycle (dc*3) for the
1985-1990 time period (tp*2) on a 30% confid. prediction (cl*2)

The above baseline vehicle implies the following data:
(the asterisk preceding a line indicates that data was
changed by the user.)

VEHICLE DATA: 2 passenger car 1 (vt*2)

Frontal area of vehicle	(af)	10.000 sqft
Coefficient of drag	(cd)	0.300
Curb weight of vehicle	(cw)	427.000 kg
Vehicle length	(vl)	9.100 ft
Power system volume	(pav)	4.300 cuft
Power system weight	(paw)	109.000 kg
Power-to-weight ratio	(pwr)	0.026
Rated range	(rd)	120.600 km
Efficiency of mtr/controller	(etmc)	0.030
Weight propagation factor	(wpl)	1.300
Motor weight to power ratio	(wmp)	4.300 kg/kw
Motor volume to power ratio	(vmp)	0.040 cuft/kw
Weight of payload	(wp)	136.000 kg

INPUT FROM MODEL

With the above data, the following design parameters have been calculated:

<u>Battery design parameters</u>	
Range (dc*3) interrupt	118.361 = km
Specific charge	30.700 = h/kg
<u>Vehicle design parameters</u>	
Payl. wt	939.34 kg
payload	136.000 kg
total test wt	1000.34 kg
vehicle length	10.11 ft
battery wt	304.10 kg
motor + controller wt	127.40 kg
transmission wt	11.89 kg
road load energy per cycle	1061.600 = h/cycle

This page of output is held in file named event
Could you like to rerun this model with more changes?
(Type in "yes", or "no")

An example for the second class of models is the EXXON econometric model for electric cars. It is available on TIS, but its prompting is that originally devised by EXXON. Here TIS simply acts as the controller for the model run and provides a convenient means of initiation and execution. Any model which can be compiled and processed on the PDP-11/70 machine can be integrated into a TIS database in this manner.

The third type of modeling capability on TIS is equally powerful. Here the model is executed on a foreign host computer under TIS control. TIS connects an authorized user to the distant computer automatically and activates the named model. An example is the Electric Vehicle Model (EVEVEC) at Jet Propulsion Laboratory.⁷² Another example is the "CCC" Thermal Aquifer Model⁷³ under development at Lawrence Berkeley Laboratory (LBL) for the Seasonal Thermal Energy Storage Program. "CCC" was moved from LBL to the Solar Energy Research Institute (SERI) computer to become part of their solar energy model library. It requires a CDC-7600 and considerable time to execute. In this case, TIS is used to prepare the input file for "CCC" execution at SERI and the retention and analysis of results.

This technique provides a very powerful capability for the execution of models at any site under TIS control. It also offers the possibility of preparing input from common data files, executing the models at different sites under TIS control, and comparison of calculated results on TIS to establish the relative accuracy of models. (Candidates for this effective procedure are the national energy models, which can not readily be moved from their home base.) With this approach, the user can also avail himself of any improved version of the model when it becomes available. Significantly, the models do not have to be translated for use elsewhere or divorced from the creative work of their originators. We are looking forward to an opportunity of using TIS in this capacity.

The output from large models can be voluminous and difficult to handle. As a remedy, we envision the extraction of significant calculated parameters from the resulting large output file for viewing and decision making on TIS. The bulk of the output file could be left at the remote host machine. TIS would maintain the index and the bookkeeping. Alternatively, as being planned for the "CCC" model, the output file could be transferred to the LBL computers under TIS control for post-processing on LBL machines and printing on their peripherals. The effective transfer of such large volumes of data requires high-speed communications with transmission rates of at least 9,600 bps.

Modeling requires programming. The major languages available on TIS today are:

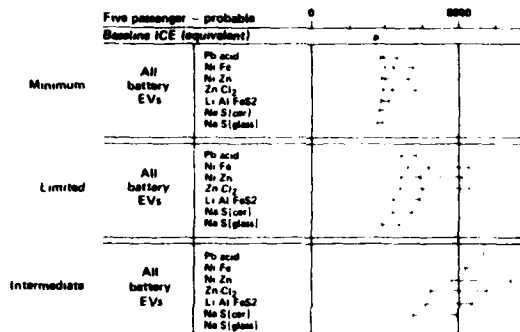
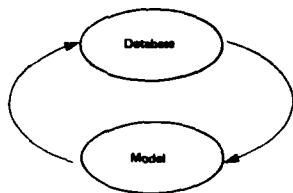
FORTRAN IV
PASCAL
MACRO II
BASIC
SNOBOL
LISP

APL
"C"
RATFOR
DC
MB
AS

Several powerful text editors are supported by the UNIX operating system and provide online editing capabilities for a variety of different terminals. Large programs, requiring extensive computer time, can be scheduled by TIS for compilation or execution at night. The results of calculations can be saved in a user-specified library of data files.

MODELING ON "TIS", CONT'D:

- Parameter studies, self-documenting
- Graphical trend analysis
- Historical record of results



Projected initial cost (1977\$) of energy-storage automobiles with reduced performance designed to maximize the use of electricity.

Several statistical and graphical analysis routines are available on TIS. We have established, especially in graphics, a number of powerful programs, some of which permit online input in a prompting manner. This pertains to the creation of bar charts, pie charts, and milestone charts for administrative purposes. The graphs can be prepared in color as hard copy or directly as viewgraphs. Once created and named, the resulting format file can be released for use by others elsewhere and printed near-instantaneously cross-country on compatible equipment.

"Communications"

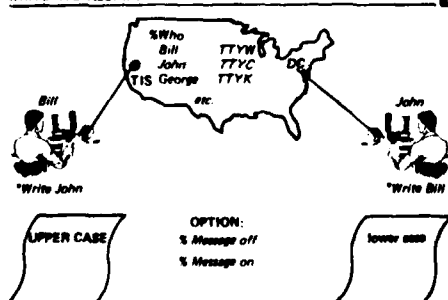
Effective communications are essential for information transfer among co-workers in different time zones. TIS offers the following capabilities:

- | | |
|--------------------------------------|--|
| comment | - a self-prompting routine to send public messages to the TIS database administrator. |
| write | - a diascript between two users. |
| link | - provides tutorials for one or a group of users. |
| electronic mail | - serves the entire user community, inclusive of voting, and the joint preparation of reports. |
| interconnection with word processors | - permits the transmission of letters and reports via TIS. |

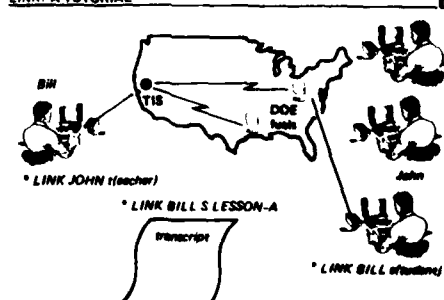
Work on an integrated system like TIS invites comments and requests for improvement to the TIS administrator. The self-guided comment routine offers this capability. These suggestions can be viewed by all users of the system and offer an opportunity to see the requests made by others, and the corrective response by TIS management.

The write command is a diascript between two users logged in on TIS at the same time. The write command, followed by the name of the user you wish to reach, prints an alert message on the remote terminal. A similar confirmation from the other side is required to establish communication. Typing then takes the place of a dialogue. A signal can be typed to indicate the end of a question or statement, inviting the response, and so forth. (If you should be doing serious work on an editor or in graphics, and do not wish to be interrupted, you can issue the message off command to silence any disturbance.)

WRITE: A DIASCRIP



LINK: A TUTORIAL

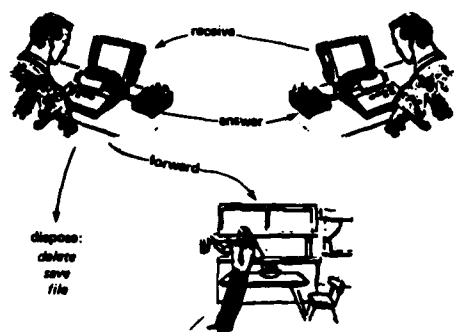


The link command is used for tutorial purposes. By previous agreement, it permits any two users to work together. One user becomes the teacher and works in the student's account. A dropfile can be created for subsequent perusal of the joint transactions. This capability is being used by TIS staff for cross-country tutorials. They are especially effective when used with a voice phone, permitting the student to see and hear instructions simultaneously. Arrangements can be made for class tutorials.

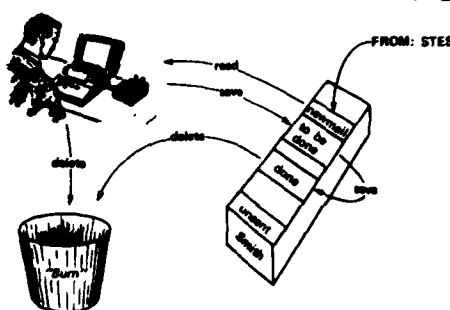
Electronic mail (em) permits you to send and receive messages, to answer and forward mail, to issue group mailings, and to file correspondence in a mail filing system of your own.⁷⁴

Some 26 different options are available to compose and edit messages and reports, correct spelling by reference to the online Webster's dictionary, send blind copies, and check whether an addressee may have already read your mail. Of course, you can delete all mail.

MAJOR OPTIONS OF em:



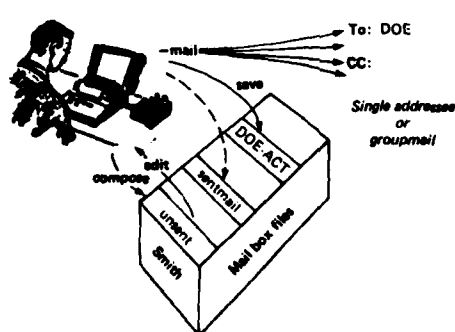
READING AND DISPOSING:



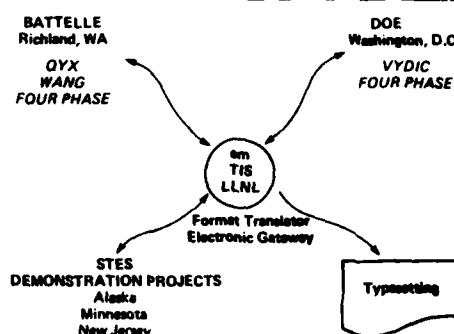
Online help is available for all options. Commands can be executed by their imperative or starting letter. This creates a very user-friendly work environment for beginners and experts.

Interconnection with Text Processors. We established the capability to connect TIS with several word processors: WANG, LEXTRON, and QYX. Connections to the FOUR PHASE system and VYDEC are planned. When used in conjunction with electronic mail, any letter or report typed on a word processor can be made part of an electronic mail message and sent on its way to the destination ahead of any written confirmations. Incompatible control characters among some of the different word processor systems are translated by TIS as required.

COMPOSING AND MAILING:



INTEGRATION WITH WORD PROCESSORS:



"Distributed Networking"

This is a very powerful TIS capability. It permits connection and use of other information centers and computers in an automated and controlled manner. At the present time, we have provisions for access to 22 other centers, thus vastly multiplying the information content and capabilities of TIS. The Network Access Machine software (NAM) used on TIS for this gateway function stems from earlier work by the NBS Computer Institute for Science and Technology.^{75,76}

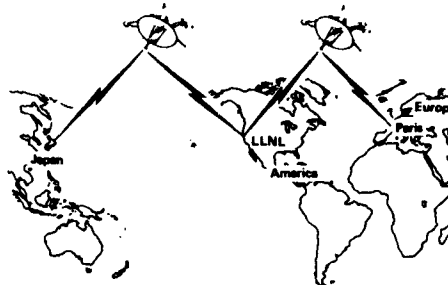
To establish a connection, the arrangements require only one contract with TIS. Individual users on TIS are then granted access as needed for the duration of their work. Audit files keep accurate records of all transactions for accounting purposes.

NETWORKING

Hampel: * Grant system DOE/RECON bill



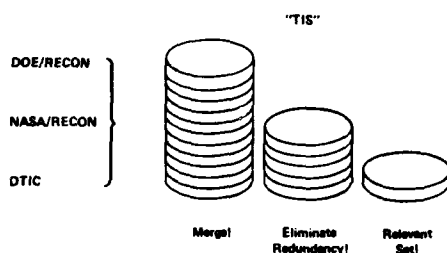
SELECTIVE SHARING OF DATA



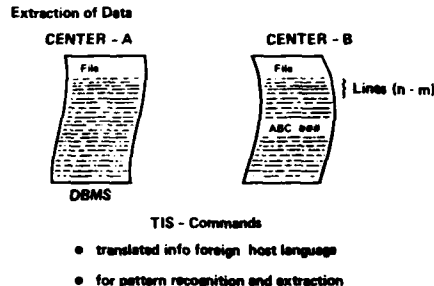
The significant aspect of this capability is that individual users of TIS need not learn the access protocols, passwords, or peculiarities of the foreign host computers. They simply select the information center by Option Number or by target name. TIS does the rest. The power of this gateway approach was recently demonstrated during the 7th international CODATA conference in Kyoto, Japan. With a simple command, "connect DARC," TIS established a computer communication from Kyoto, Japan, via TIS at LLNL, to the DARC system at the Institute of Topology at the University of Paris, France.⁷⁷ For extended periods of time each day, interactive graphics of the sophisticated DARC system were demonstrated in Japan, in full duplex at 1200 baud, without any noticeable delay in transmission.

TIS is also capable of retaining the viewed or extracted information, derived from a foreign host, in the user's account for subsequent processing and use where legally permissible. A cogent example is our interconnection to the extensive DOE/RECON information system. All citations retrieved can be placed into a file, aggregated, and processed interactively online for the immediate creation of subject and author indexes, or for topical concordances and text analysis in general. Any bibliographic field element can be cross-correlated with any other. Where required, citations can be complemented with key-to-disk annotations about their relevancy and ranking. Requests for full-text copies can be issued automatically. Citations can be augmented with numeric or descriptive data derived from the reports. Publication statistics can be shown graphically, online, and enhance further the insight possible from the retrieved information. This opens new vistas for extraction of higher intelligence from descriptive text in science and technology.

COMPILATION FROM DIFFERENT SOURCES:

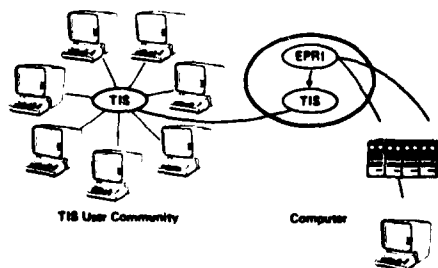


POST-PROCESSING OF NUMERIC INFORMATION



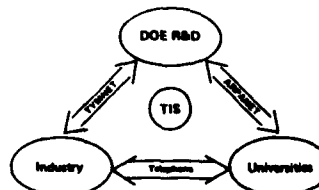
We expect to have similar links soon with NASA/RECON and with the unclassified Defense Technical Information Center (DTIC). These files are in the public domain and could be used to establish comprehensive, well-indexed bibliographies, now carried out more laboriously. This capability is equally applicable to numeric data and offers the opportunity of data aggregation from different sources into one topical summary. Use of commercial systems in this manner requires careful study of legal aspects and contractual agreements. With regard to the transfer of technology, we are making good use of the following procedure which does not require transfer of the data or models to TIS, and retains full control in the hands of the originators: A computer account is opened for the TIS user community on the remote host computer where the resources are located. The owners of the data and models release, periodically, those versions which they are prepared to share with TIS.

TIS INFORMATION TRANSFER



NO TRANSLATION! ORIGINAL DATA AND MODELS ARE USED!

TIS - A GATEWAY FOR TECHNOLOGY TRANSFER



CONCLUSION

I believe we can look forward to the 1980s with excitement. In the past, technology often lagged behind our information requirements. The reverse may soon be true. We probably find it difficult to imagine ourselves using a pocket-portable flat-screen CRT terminal, homed in on a satellite and capable of searching and reading -- wherever we are -- any of the 5.2 million retrospective MARC files of the Library of Congress now being installed on DIALOG, or conducting a molecular structure search on the 4-million-large Structure and Nomenclature Search System, soon to be offered by CAS and CIS. On the home front, the new 9-digit ZIP code of the U.S. Postal Service should be capable, by itself, of delivering mail to anyone in the United States directly by cross-correlation with computer-based address lists and/or social security numbers. Orwell's 1984 is only three years off.

The "Network Nation" is more than the title of a recent book.⁷⁸ It marks the beginning of a new decade in which intelligent factual information may become the scarce and costly resource. National and international organizations are trying to come to grips with this situation. CODATA's emphasis is on data in science and technology.⁷⁹ UNESCO recognizes basic requirements for factual data in developing countries.⁸⁰ Regional information centers in the Far East and Africa are being planned that could transfer the know-how of the post-industrial countries to those anxious to learn, but foresighted enough not to repeat the mistakes of the past. There are also signs of concern, one of which is the abuse of information. We are learning that CIS is bringing online for worldwide access commercially nonconfidential data and production volumes of chemical manufacturing plants in the United States.⁸¹ We are also aware that some countries find it expedient and in their interest to model the economy of the United States with U.S. demographic and time-series data on U.S. computers.

In our democracy, free access to information is essential. We should not be remiss in using it ourselves, first!

May God grant us the wisdom to know how to proceed.⁸²

ACKNOWLEDGEMENT

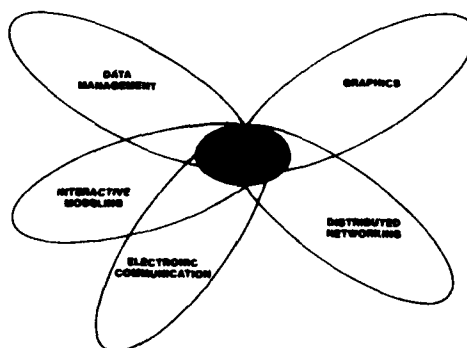
Robert Tap, Director of the Information Division of the Transportation Systems Center for the Department of Transportation (DOT-TSC), specified in 1975 the functional requirements for a transaction controller that could serve as an automated gateway for DOT system analysts. This resulted in the Monitor of Distributed Data Systems (MODDS) and laid the foundation for the work later continued by the Department of Energy.⁸³

The Technical Information System (TIS) owes its beginning to the foresight and technical judgement of George F. Pezdirtz, then Director of the DOE Division of Energy Storage. He recognized years in advance that information science and communications technology were on converging paths, and that this merger could be directed toward faster and better access to information communications among administrators and R&D personnel of an important national program, and result in increased productivity. This organization (Office of Energy Systems R&D) has been the prime sponsor of TIS.

The staff of Control Data Corporation and of Schriebman Consulting have contributed unique expertise and talents to the project.

Lawrence G. O'Connell, Manager of the Transportation Systems Research program at LLNL, fostered the development of the Technology Information System. It is used as a practical and useful tool for cross-country information management in support of electric vehicle research.

This report was expertly typed and designed by our Word Processor Specialist Ellen Gerhard, with assistance from Carol Addison, and Donna Seacutto.



REFERENCES

1. The Conference Board: Information Technology - Some Critical Implications for Decision Makers. Published periodically by the Conference Board, Inc., 845 Third Avenue, New York, NY, 10022, 1972.
2. IBM, Information Systems Planning, 5-day Seminars given on the East and West coast by the IBM Systems Science Institute, International Business Machines Corporation, Data Processing Division, 1133 Westchester Avenue, White Plains, NY 10604; IBM World Trade Corporation, 360 Hamilton Avenue, White Plains, NY 10601.
3. Heller, S. R., Economics of Online Data Dissemination, 7th International CODATA Conference, Kyoto, Japan, 1980.
4. Abelson, P. H., Use of Data in Basic and Applied Science, Keynote address, 7th International CODATA Conference, Kyoto, Japan, 1980.
5. The present major computer and communication facilities at Lawrence Livermore National Laboratory, located 45 minutes East of San Francisco, on a 2 mile square area are:

Computers:	4 - CDC-7600
	4 - Cray-1
	Numerous mini-computers
Communications:	MFENET
	SACNET
	2 - Satellite Stations (WESTAR-2 & 3)
	4 - ARPANET
	16 - TELENET/TYMNET
	360 - Dedicated lines
	7000 - Telephones
Word Processors:	160 - Word Processors (WP) CPUs
	650 - WP Work Stations
6. Blumberg, Stanley A., Energy and Conflict. The Life and Times of Edward Teller, G. P. Putnam's Sons, New York, 1976, pp. 95-96.
7. Shimanouchi, T., Two Projects on the Evaluation and Compilation of Scientific Data in Japan, Proceedings of the Fifth Biennial International CODATA Conference, Boulder, Colorado, June 28-July 1, 1976.
8. A Report of the Science Council, A New Plan for a Science Information System in Japan, The Ministry of Education, Science and Culture, January, 1980.
9. Matthoefer, H., The Third Data Processing Program of the Federal Republic of Germany, for 1976-79, published by the Directorate for Research and Technology, Bonn, Bonn University Press, ISBN-3-88135-007-1, 1976.
10. Euronet DIANE Information, Jean Monnet Building, B4 009 CEC, Luxembourg, (Grand Duchy).
11. Committee on Data Needs, Numerical Data Advisory Board, Assembly of Mathematical and Physical Sciences, National Academy Of Sciences, National Needs for Critically Evaluated Physical and Chemical Data, 1978.
12. Personal communications with Dr. George Michael, Lawrence Livermore National Laboratory, 1981.
13. Michael, G. A., Mass Archival Storage: Some Trends, Needs and Plans at DOE Laboratories, Lawrence Livermore Laboratory, UCRL-82354 Rev. 1, 1979.
14. Michael, G. A., Archival Storage Requirements, Lawrence Livermore Laboratory, UCRL-82513, 1979.
15. Buzbee, B. L., Worlton, W. J., Michael, G., and Rodrigue, G., DOE Research in Utilization of High-Performance Computers, Los Alamos National Laboratory and Lawrence Livermore National Laboratory, LA-8609-MS, 1980.
16. Landau, R. N., Wanger, J., and Berger, M. C., Directory of Online Databases (Vol. 1, No. 3) Cuadra Associates, Inc., 1523 Sixth Street, Suite 12, Santa Monica, CA 90401, 1980.
17. Williams, M. E., Lannon, L., O'Donnell, R. and Barth, A. N., Computer-Readable Data Bases, A Directory and Sourcebook. American Society for Information Science (ASIS), 1979.
18. Learned Information, Eusidic Database Guide 1980, Besselsleigh Road, Abingdon Oxford OX13 6EF, England, ISBN 0 904933 20 2.
19. Westbrook, J. H. and Desai, J. D., Data Sources For Materials Scientists and Engineers, General Electric, Schenectady, New York, Report No. 78CRD046, 1978.
Refer also to: Kruzas, A. T., Encyclopedia of Information Systems and Services, Third Edition. Gale Research Co., Book Tower, Detroit, Michigan, 48226, 1978.
20. Fry, J. P., and Sibley, E. H., The Development of Data-Base Technology, ACM Computing Surveys, Special Issue: Data Base Management Systems, Vol. 8, No. 1, 1976.
21. Hampel, V. E., Database Management Systems for Numeric/Structured Data. 7th International CODATA Conference, Kyoto, Japan, October 1980. UCRL-84594 & Proceedings of Conference.
22. Auerbach Editorial Staff, Data Base Management. Auerbach Publishers Inc., 6560 N. Park Drive, Pennsauken, NJ 08109, 1980.
23. Datapro, A Buyer's Guide to Data Base Management Systems. Datapro Feature Report, Delran, New Jersey, 1980.
24. CODASYL Systems Committee. Selection and Acquisition of Data Base Management Systems. New York, NY: Association for Computing Machinery (ACM); 252 p. ACM Order Dept., P. O. Box 12105, Church Street Station, New York, NY 10249, 1976.
25. CODASYL Systems Committee, Feature Analysis of Generalized Data Base Management Systems, CODASYL Systems Committee Technical Report, May 1977.
26. Seymour, J., Fife, D., Deutsch, D., and Sockut, G., Architectural Considerations For Federal Database Standards, Institute for Computer Sciences and Technology, National Bureau of Standards, Proceedings of 18th IEEE Computer Society International Conference, San Francisco, CA, 1979.
27. Tubbs, N., Generalized Data Management Systems and Scientific Information, OECD Nuclear Energy Agency, U.S. Dept. of Energy, U.S. National Bureau of Standards, 38 bd. Suchet, 75016 Paris, France. CONF-771030, 1978.
28. NASA, Engineering and Scientific Data Management. Proceedings of a Conference held at Langley Research Center, Hampton, VA, May 18-19, 1978. ICASE. NASA Conference Publication 2055, 1978.
29. Klingbiel, P. H., A Technique For Machine-Aided Indexing, Defense Supply Agency, Defense Documentation Center, Inform. Stor. Retr. Vol. 9, pp. 477-494. Pergamon Press, 1973.
30. Klingbiel, P. H. and Rinker, C. C., Evaluation of Machine-Aided Indexing, Defense Documentation Center, Cameron Station, Information Processing & Management, Vol. 12, pp. 351-366. Pergamon Press, 1978.

31. Lerner, R. G., Feinman, R., and Lieblich, I., Data-Descriptive Records in the Physical Sciences, American Institute of Physics, 335 East 45th Street, New York, NY 10017, AIP-76-1, 1976.
32. Hampel, V. E., Wiley, R. A., Integrating Text and Data for Energy Research, Lawrence Livermore National Laboratory and Los Alamos Scientific Laboratory, UCRL-76991; also published in the Proceedings of the ERDA-wide Conference on Computer Support of Environmental Science and Analysis, Albuquerque, NM CONF-750706, July 9-11, 1975.
33. Williams, M. E., Annual Review of Information Science and Technology. American Society for Information Science (ASIS), 1979.
34. Organized each year by Online Review, The International Journal of Online & Teletext Information Systems, Online Information, Learned Information, Oxford and New York.
35. Annual Conferences on Very Large Databases.
36. Williams, M. E., Research Toward a Transparent Information System, Information Retrieval Research Laboratory, Proceedings of Perspectives on Scientific and Technical Information, Oak Ridge, Tenn, ORNL-5749, 1980.
37. Killeen, J. and staff, The National MFE Computer Center: Increased Capability in 1978, Energy and Technology Review, Lawrence Livermore National Laboratory, September 1978.
38. Hampel, V. E., and Henry, E., The Time for Atomic and Molecular Databases is Now!, Lawrence Livermore National Laboratory, UCRL-79286, 1977.
39. Hilsenrath, J., Summary of On-Line or Interactive Physico-Chemical Numerical Data Systems, Office of Standard Reference Data, National Bureau of Standards, Washington, D.C., NBS Technical Note 1022, 1980.
40. Hall, D., Scherrer, D., Sventek, J., A Virtual Operating System, Communications of the ACM, September 1980.
41. Donnelley, J. E., Managing Domains in a Network Operating System, Lawrence Livermore National Laboratory, UCRL-85562, Proc. of the International Conference on Local Networks & Distributed Office Systems, London, May 11-13, 1981.
42. General Purpose Computer Networks and Resource Sharing in ERDA, Volume 1, 2, 3, Summary, Conclusions and Recommendations, Working Group on Computer Networking, Division of Basic Energy Sciences, ERDA, July 15, 1977.
43. Proceedings of the Berkeley Conferences on Distributed Data Management, and Computer Networks, CONF-790834, 1979 and 1980.
44. American Society for Testing and Materials, Standard for Metric Practice, 1916 Race Street, Philadelphia, PA 19103.
45. Hampel, V. E. and Ries, D., Requirements for the Design of a Scientific Data Base Management System, Lawrence Livermore Laboratory, UCRL-80171, Rev. 1, Proceedings on Generalized Data Management and Scientific Information, sponsored by the OECD Nuclear Energy Agency, U.S. DOE, & National Bureau of Standards, Lawrence Berkeley Laboratory, CONF-771030, 1977.
46. David, A. and editor, Engineer's Needs for Scientific Technical Information, World Federation of Engineering Organizations, Committee on Engineering Information, Paris, 1979.
47. Selander, S. E., A Syntax for Machine-Readable Description of Numeric Data, Mitre Corporation, 1979.
48. Winter, R., Lozano-Perez, T., Eastlake III, D. E., and Rosenberg, R. L., Overview of the Chemical Substances Information Network, Computer Corporation of America, 575 Technology Square Cambridge, MA, CCA-81-03, 1981.
49. Lide, D., Critical Data for Critical Needs, Science, Vol. 212, No. 4501, 1981.
50. The National Commission on New Technological Uses of Copyrighted Works, Final Report, Library of Congress, 1978.
51. Bush, G. P., Dreyfuss, R. H., and editors, Technology and Copyright: Sources and Materials, 2nd Edition, Lomond Publications, Inc., Airy, MD, ISBN 0-912338-17-2, 1979.
52. Uhlig, R. P., Farber, D. J., Bair, J. H., The Office of the Future, Communications and Computers, North-Holland Publishing Company, 1979.
Refer also to: National Research Council, Board on Telecommunications-Computer Applications, Toward the Electronic Office, Washington, DC, Guidance on Requirements Analysis for Office Automation Systems, NBS Special Publication 500-72., U.S. Department of Commerce, 1980.
53. Woolfe, R., VIDEOTEX - The New Television/Telephone Information Services, Heyden, 1980.
Refer also to: Marsh, F. E., VIDEODISC Technology, DOE Computer Operations Branch, Technical Information Center, CONF-810322-1, 1981.
Refer also to: Hirsch, VIDEOTEX in the U.S., Computer World, Washington, DC Bureau.
NOTE: Viewdata and Online Systems and The Potential Impact of Viewdata and Online Retrieval Systems 1980-2000 are some of the first courses offered in this exciting field by H. Leonard Fisher from Lawrence Livermore National Laboratory, as the instructor, at Golden Gate University, San Francisco, CA.
54. Kurzweil, Computer Products, Inc., 33 Cambridge Parkway, Cambridge, MA 02142.
55. Lempel, A., Cryptology in Transition, Computing Surveys, Vol. 11, No. 4, 1979.
Refer also to: National Bureau of Standards, Federal Information Processing Standards, Publication 46, 1977.
Refer also to: Diffie, W. and Hellman, M. E., New Directions in Cryptography, IEEE Trans. Inform. Theory IT-22, pp. 644-654, 1976.
56. Hampel, V., McGrogan, S. K., Gallo, L. E., and Swanson, J. E., THE LLL "META-MACHINE" -- A Flexible, Extensible, and Practical Technique for Interactive Data Management, Modeling, and Distributed Networking, Lawrence Livermore National Laboratory, Livermore, CA, UCRL-83064, May 1979.
Refer also to: Hampel, V. E., et al, The META-MACHINE - a New Way of Making a Mini-Computer Do a Maxi-Job in Data Management, Lawrence Livermore National Laboratory, UCRL-81170, Presentation to the CUBE Symposium, October 4-5, Los Alamos, NM, 1978.
57. Woodfill, J. et al., INGRES, Version 6.2, Reference Manual, Electronics Research Laboratory, College of Engineering, University of California, Berkeley 94720, 1979.
Refer also to: Epstein, R., Creating and Maintaining a Database Using INGRES, NCB/ERL M77/71, 1977, and to the review article on other current relational systems by Rita Shoor in the February 9th, 1981, issue of "Computerworld."
Refer also to: Schneidawin, J., Users of Bell's UNIX Seen Spending \$9.58 Billion by 1986, Management Information Systems, p. 17, July 22, 1981.
58. The Intelligent Database Machine (IDM-500), Britton Lee, Inc., Cupertino, CA.
Refer also to: The excellent review article on other database machines by Vincent C. Rauzino in the December 29, 1980/January 5, 1981 issue of "Computerworld."

59. Koponen, B. L., Wilcox, T. P., and Hampel, V. E., Nuclear Criticality Experiments from 1943 to 1978, An Annotated Bibliography in Three Volumes, Lawrence Livermore National Laboratory, UCRL-52769, 1979.
60. American Association for the Advancement of Science, SCIENCE, Vol. 211, No. 4483, February 13, 1981.
Refer also to: Lawrence Berkeley Laboratory, NEWS MAGAZINE, Vol. 5, No. 2, Summer, 1980.
One of the best known products developed by the LLNL Graphics Group is the portable, generalized, color graphics program: Rowe, J. Keller, P., and O'Hair, K., GRAFCORE Reference Manual, Lawrence Livermore National Laboratory, LCSD-434, 1980.
61. Vickers, D. L., Smith, G. S., and Levine, S. R., Lawrence Livermore National Laboratory, and L. G. Cross, The Multiplex Company, San Francisco, California, Two Unusual Three-Dimensional Displays, Trans. Am. Crystallographic Assn., Vol. 12, 1976.
62. Vickers, D. L., The Integral Hologram as a Scientific Tool, UCID-17035 Lawrence Livermore National Laboratory, California, 1976.
63. Jeong, T. H. and Snyder, H., SID Int'l. Symp., Digest of Technical Papers 5, 1974.
64. Sutherland, I. E., Proc. Fall Joint Computer Conference 33, 757, 1968.
65. Hampel, V. E., Gallo, L. E., Kavin, R. A., Kopytoff, V., McGrogan, S., O'Connell, L. G., Pavel, G., Schriebman, J. A., Swanson, J. E., Editor, Proceedings of the First DOE/LLL Workshop on the Technical Management Information System (TMIS), CONF-7912-8, 1979.
66. Hampel, V. E., Gallo, L. E., Kavin, R. A., Kopytoff, V., Mallon, B., McGrogan, S., Pavel, G., Rabe, W. G., Schriebman, J. A., Scott, W. S., and Swanson, J. E., User's Manual for the Technology Information System, 2nd Edition, M-0112, 1981.
Refer also to: Hampel, V. E., Decision Making with Interactive Access to Integrated Administrative and Technological Databases, Lawrence Livermore National Laboratory, UCRL-80353, Proceedings of the ERDA/AESOP XVII Conference, Boston, MA, September 13-15, 1977.
67. The Transportation System Research program at LLNL conducts theoretical and experimental research for future electric vehicles, the innovative aluminum-air battery, roadway powered electric vehicles, route analysis for the U.S. Postal Service., and transportation-related R&D in general. The computational analysis and performance predictions are made on the Technology Information System. Some are cited here to convey the complexity of calculational-graphics, and the extent of information & data management carried out on the TIS intelligent gateway computer:
 - a) O'Connell, L. G., et al., Energy-Storage Systems for Automobile Propulsion: Final Report, Volumes 1 - 4, Lawrence Livermore National Laboratory, Livermore, CA, UCRL-53053-80, 1980.
 - b) Behrin, E. and Cooper, J. F., Aluminum-Air Batteries for Vehicles, Lawrence Livermore National Laboratory, Livermore, CA, UCRL-85605, 1981.
 - c) Walter, C. E. and Salisbury, J. D., Inductive Energy Transfer in Transportation, Lawrence Livermore National Laboratory, Livermore, CA, UCRL-85141, 1980.
 - d) O'Connell, L. G., A Comparison of Energy-Storage Devices for Use in Future Automobiles, Lawrence Livermore National Laboratory, Livermore, CA, UCRL-85654, 1981.
68. Simmons, W. W., The NOVA Laser Fusion Facility, Lawrence Livermore National Laboratory, Energy and Technology Review, 1980.
69. Janz, G. J., Allen, C. B., Downey, Jr., J. R., and Tomkins, R. P. T., Physical Properties Data Compilations Relevant to Energy Storage -- Molten Salts: Eutectic Data, Molten Salts Data Center, Cogswell Laboratory, Rensselaer Polytechnic Institute, Troy, NY, NSRDS-NBS 61, Part 1, March 1978.
70. Whisler, D. J., Behrin, E., and Salisbury, J. D., Aluminum-Air Vehicle Simulation Programs User's Manual and Documentation, Lawrence Livermore National Laboratory, Livermore, CA, M-113, 1980.
71. Kong, M. K. and Walter, C. E., Route Profile Analysis System Report Covering Task A - CARA Model Verification, Lawrence Livermore National Laboratory, Livermore, CA, UCID 19027, 1981.
72. Chapman, P., Hybrid Vehicle Simulation Program (ELVEC Version 6.0); Volume I: Program Documentation, Jet Propulsion Laboratory, 5030-247, Rev. D, 1981.
73. Mangold, D. C., Lippmann, M. J., and Bodvarsson, G. S., Computer Program for the Numerical Solution of the Heat and Mass Flow Equations for a Liquid Saturated Medium - User's Manual, Lawrence Berkeley Laboratory, LBL-10909, 1980.
74. Hampel, V. E. and Schriebman, J. A., A Personal System for Electronic Mail, Lawrence Livermore National Laboratory, Manual M-115., November, 1980.
Refer also to: Schriebman, J. A. and Swanson, J. E., Crosslinking of Computer Terminals for Nationwide Tutorial Instruction, Lawrence Livermore National Laboratory, UCRL-52964, November, 1980.
Refer also to: Information Systems: Voice & Data's Mailbox Uses Any Phone, Management Information Systems, July 15, 1981.
75. Rosenthal, R. and Lucas, B. D., The Design and Implementation of the National Bureau of Standards' Network Access Machine (NAM), Institute for Computer Sciences and Technology, National Bureau of Standards, NBS Special Publication 500-35, 1978.
76. Rosenthal, R., A Review of Network Access Techniques with Case Study: The Network Access Machine, National Bureau of Standards, NBS Technical Note 917, July 1976.
77. Dubois, J. E., Analysis of Data Bank Implementation and Computer Aided Design in Chemistry, Proceedings of the Fifth Biennial International CODATA Conference, Boulder, CO, 1976.
78. Hiltz, S. R., Turoff, M., The Network Nation, Human Communication via Computer, Addison Wesley Publishing Company, 1978.
Refer also to: Fedida, S., Malik, R., Viewdata Revolution, John Wiley & Sons, 1979.
79. Watson, D. G., Rossmassler, S.A., Data Handling for Science and Technology, An Overview and Sourcebook, CODATA & UNESCO sponsored, North-Holland Publishing Co., 1980.
A corollary publication on Database Management in Science and Technology is in preparation.
80. UNISIST Newsletter, General Information Programme, Vol. 8, No. 4, 1980, and Vol. 9, No. 1, 1981.
Refer also to: Study on an International System Relating to New and Renewable Energy Sources, UNESCO, General Conference, 21 C/INF.10, Belgrade, 1980; Report on the ad hoc Expert Group on Information Flows, A/CONF.100, January 1981.
81. New Chemical Database has Non-Secret Plant Information, Announcement by NIH/EPA, July 1981.
82. The Bible, The Epistle of James, Chapter 1, Verse 5.
83. Biras, E. W., Donnelley, J. E., and Yeh, J. W., A Monitor of Distributed Data Systems (MODDS): Part-1 and 2, Digest of Functional Specifications for the Department of Transportation, UCID-17314, Lawrence Livermore National Laboratory, November 15, 1978.

DISCLAIMER

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government thereof, and shall not be used for advertising or product endorsement purposes.

Work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

**DATA
FILM**